Uncertainty Estimation in the Quantitative Interpretation of Inverted Reservoir Property Volumes

---

A Thesis Presented to

The Faculty of the Department of Earth and Atmospheric Sciences

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

By

Grant Whittle

December 2016

Uncertainty Estimation in the Quantitative Interpretation of Inverted Reservoir Property Volumes

_____

Grant Whittle

APPROVED:

_____

Dr. John Castagna, Chairman
Department of Earth and Atmospheric Sciences

_____

Dr. Evgeny Chesnokov
Department of Earth and Atmospheric Sciences

_____

Dr. Jacques Leveille
Ion Geophysical

_____

Dean, College of Natural Sciences and Mathematics

Acknowledgements

- I would like to thank Prof. John Castagna for giving me the inspiration and guidance needed to complete this thesis.
- I would like to thank Jacques Leveille and Sean Boerner for their technical guidance and support the two years at Ion Geophysical.
- I would like to thank Ion Geophysical for letting me use their BuffaloHorn Dataset for this Thesis.
- I would like to thank Shihong Chi, Howard Rael, and Scott Singleton for their prior work on the dataset, without which the thesis would not have been completed on time.
- I would like to thank my father Gary Whittle for giving me the drive to keep pushing for the best result possible in the allocated time frame.

Uncertainty Estimation in the Quantitative Interpretation of Inverted Reservoir Property Volumes

---

An Abstract of a Thesis

Presented to

The Faculty of the Department of Earth and Atmospheric Sciences

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

By

Grant Whittle

December 2016

ABSTRACT

Volumetric uncertainty in reservoir property volume estimation using 3D seismic data, well logs, and P-wave inversion outputs can be calculated using synthetic modeling, multi-attribute linear regression, and collocated cokriging. This can be accomplished using a multi-attribute linear regression to create the initial reservoir property volume, and then using this volume as a covariate to a simulated collocated cokriging approach from which an uncertainty in the volume estimate is computed. In a synthetic test example, the initial reservoir property volume estimated from the synthetic dataset exhibited a 0.92 correlation coefficient to the known reservoir properties. Due to the high correlation between the hard data and soft data the collocated cokriging output was almost identical to the multi-attribute non-linear regression. Except around the vicinity of the well were it had an overall smoother output. An uncertainty volume generated from the standard deviation of thirty realizations of the collocated cokriging process run using SGS (Sequential Gaussian Simulation) effectively predicted the lower error regimes in the vicinity of the well locations, however the areas with high error away from the well locations were not captured by this process.

When applying the framework validated above to a real dataset in the Mississippi Limestone near Morrison Oklahoma the output of the non-linear regression based method had a correlation coefficient of 0.81 to measured well logs. The collocated cokriging process created a higher vertical resolution output than the non-linear regression output because the vertical sampling is closer to that of the well, approximately 74,000,000 cells

in the grid compared to 14,000,000 in the synthetic. The overall approach shows the potential to calculate volumetric uncertainty in reservoir property volume estimation using 3D seismic data, well logs, and P-wave inversion outputs, which can be computed on a regular basis using multi-attribute linear regression and collocated cokriging.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Inspiration

The integration of seismic data and well-log properties is an important process in reservoir characterization. The main purpose of this integration is to extend the well log data, which has high vertical resolution, into the spatial volume imaged by the dense three-dimensional seismic data, which has low vertical resolution. Inverse modeling of the logs from the seismic data is used to predict certain log properties throughout the dense 3D seismic volume. Using seismic modeling or rock physics, a physically justifiable relationship between a seismic attribute and the well log property can be inferred (Kalkomey, 1997). The more constraints or a priori information included in the inversion, the more accurate the estimation can be. Thus, it has been found that use of multiple seismic attributes can help to lower the prediction error of the inverted log property. Different combinations of attributes may have a stronger correlation with the log property than others. An exhaustive search or stepwise regression must be done for every attribute combination to find the best set that correlates with the log property. A well-correlated linear relationship between a log property and a single seismically-derived attribute is possible; however more often, a linear transform of several attributes can better predict for that single log property. Multiple linear regression is one way of predicting for a log property. Non-Linear transforms of attributes can also yield better prediction results. The use of neural networks allows for an effectively higher polynomial to be fitted that may

better predict the log property. Caution must however be exercised when predicting reservoir properties from transformed attributes. One should always seek a geophysical explanation for any positive correlation between transformed attributes and measured log properties.

Ultimately, the log property prediction is only as good as the error associated with it. Understanding what this error means is very important for the interpretation of the predicted volume. Throughout the multi-attribute transform process several errors are assessed. These errors are associated with the number of attributes used in the prediction and the number of wells to be used in the prediction. A validation error is computed to optimize the performance of the prediction. Cross validation is useful in determining how many attributes can be used before there is an increase in validation error. In the process of cross validation, spurious wells can be identified and removed.

The current analysis of errors in the multi attribute transforms process assesses error only at the well locations. It would be more useful from an interpretation standpoint to view an error as a function of spatial and temporal position. This would indicate areas that should yield reliable interpretations of the reservoir property and those which are not predicted as confidently. Extrapolating this error spatially across the predicted volume would help in the overall quantitative interpretation of the prediction. The current non-linear regression technique used in this research is unable to produce such a spatially and temporally variable uncertainty volume. The application of cokriging and sequential Gaussian simulations, a geostatistical interpolation technique, will be used to create this uncertainty volume.

1.2 Project Overview

Multi-attribute prediction and uncertainty analysis will be applied to a real world dataset as well as a synthetic dataset. The area of the real world dataset is in Oklahoma located in the counties of Noble, Pawnee and Payne. The geologic province map, shown below in figure 1, shows where the seismic SCAN dataset is located in Oklahoma.

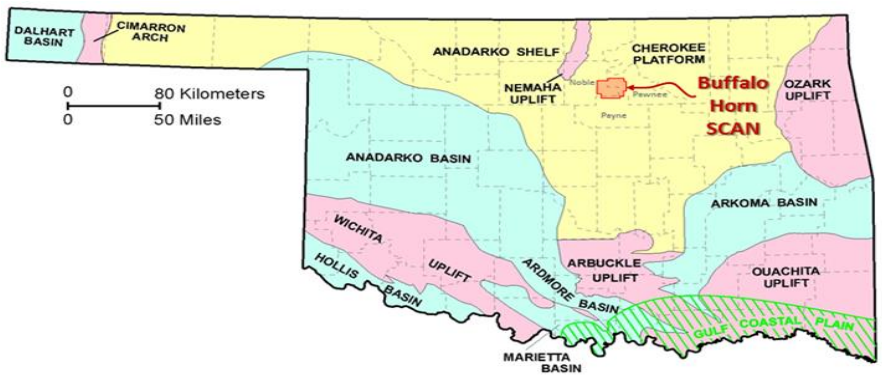


Figure 1: Geologic Province Map and Boundary of Dataset *(Johnson, 2008)*

The reservoirs of interest in the area are the Mississippi Limestone and Woodford shale. There is a dense distribution of wells in the area of the 3D seismic. In this research 13 of these wells will be used. The actual dimensions and details of the dataset can be found in an upcoming chapter. Figure 2 below is a simple map view of the boundaries of the seismic dataset and the wells used. Since the data used for this research includes propriety well data and seismic data, all maps will be shown at a relative scale. The cross sections in the paper are all displayed in time.

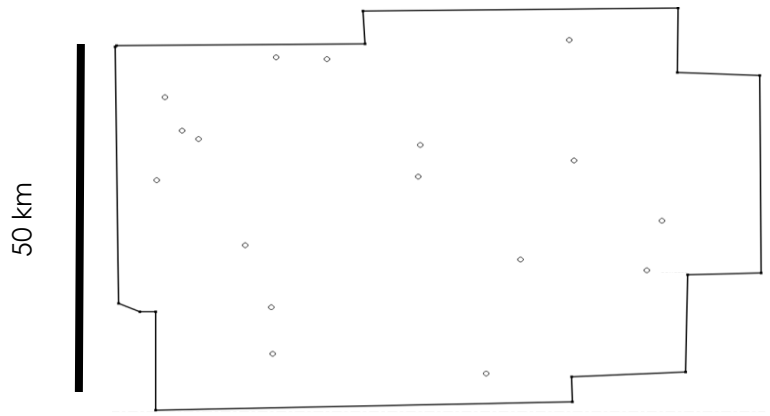


Figure 2: 3D Seismic Boundary and Well Locations

The purpose of this research is to understand uncertainty in the prediction of reservoir attributes using multiple attributes. To be more specific, this research focuses on spatial uncertainty. Thus, a process will be created to generate an uncertainty volume of the predicted reservoir volume. To verify how accurate the process is, it will be tested against a dataset where the uncertainty is known, or the reservoir property volume predicted is known. The synthetic dataset to be created will be of similar characteristics to the real 3D seismic dataset. To achieve this, a synthetic dataset will be created from the inversion outputs of the real dataset. Figure 3 shows a workflow of the entire process. After an initial elastic inversion it is applied symmetrically to both the synthetic dataset and real (BuffaloHorn SCAN) dataset.

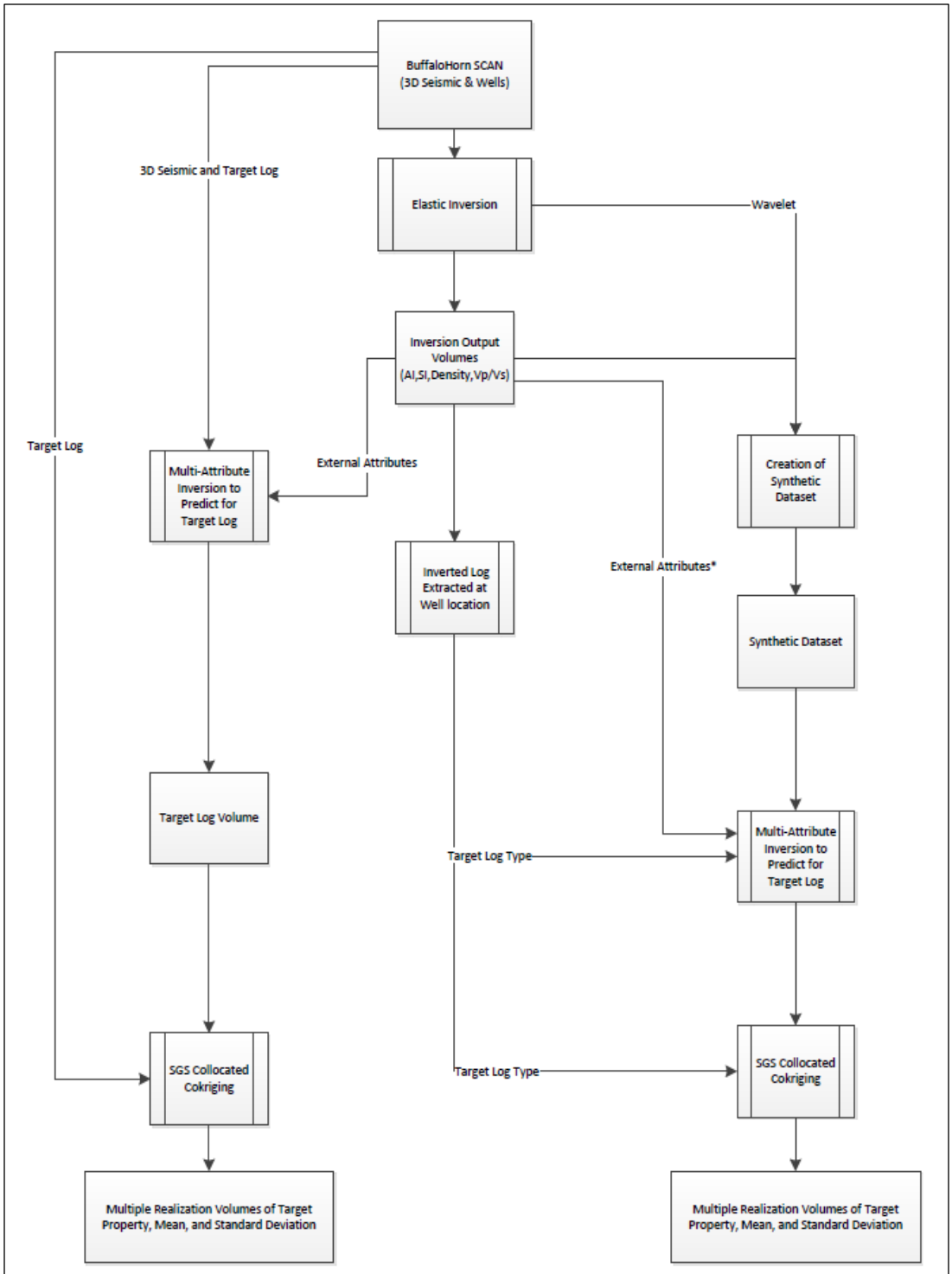


Figure 3: Project Workflow

1.3 Problem Statement

A major problem in characterizing a reservoir is the lack of data that can be trusted. Well data is deemed "ground truth" in comparison to other data such as seismic, gravity, magnetics, etc. However, the area of interest for most exploration and development operations is sparsely populated with well data. The geologist or geophysicist is then tasked to interpret the areas between and around the wells. This interpretation can be aided by other data such as 3D seismic; however, the seismic volume does not measure the same properties as the well data, nor approach their vertical resolution. The well data has multiple log suites that inform us about rock property information such as porosity and density. The well data also has much higher frequency content than that of seismic data.

Yet the seismic volume has the ability to guide the interpretation of the reservoir between and around wells. There are attributes derived from the seismic data that correlate well with some of the log properties. Rarely is a single seismic attribute linearly related to the log properties of choice, yet multiple seismic attributes can be linearly transformed to correlate well with a log property. Even better, a nonlinear transform of the multiple seismic attributes may improve fit to the log property. The geostatistical method of multi-linear transforms does just that. This method has been around for the past couple of decades. However, there are inherent issues with this method that are rarely discussed.

The purpose of this paper is to describe these issues and show them in a way that will help further interpretation of the reservoir attribute. The underlying issue with using multi-attribute transforms is defining the spatial uncertainty of the transform.

"The extrapolation uncertainty must also be addressed, i.e., the behavior of the prediction between and away from the training wells. Leonard suggests an approach to determine if there is sufficient training data by estimating local training data density in the neighborhood of the input sample to be used for the prediction." (Hampson, Schuelke, & Quirein, 2001)

It is important to note here that other geostatistical methods predict for a log property using multiple variables and give the spatial uncertainty. Cokriging is one such method. A comparison of multi-attribute linear regression and cokriging will be shown in this paper but that is not the main purpose of the research. Before venturing forward into how to describe the spatial uncertainty, one must describe the method of multiple attribute linear regression used in this research.

Chapter 2 Multivariate Non-Linear Regression

2.1 Prediction of Rock Properties with Seismic Attributes

Before predicting for a chosen target log property with a seismic attribute a valid relationship between them must be determined. One way to investigate this relationship is to cross-plot the two. In doing this one assumes that the target log has been converted from depth to travel-time at the same sample rate as the seismic attribute. Further smoothing reduces the resolution of the target log to the same resolution of the seismic attribute. The

target log usually has a much higher sample rate than that of the seismic attribute. Once the cross plot has been made, one can then visually investigate whether a linear relationship exists between the seismic attribute and the log property. The strength of this linear relationship may be quantified by a simple regression as described below.

$$y = a + bx \tag{2.1}$$

In Equation 2.1, y is the target log property and x is the attribute to predict with. The coefficients a and b in equation 2.1 can be computed by minimizing the mean squared prediction error (Hampson, Schuelke, & Quirein, 2001). Equation 2.2 describes the mean squared prediction error, E. The prediction error is a measure of goodness-of-fit for the regression line. The linear relationship can be relaxed by using non-linear transforms on the single seismic attribute or well log property, which has been shown in some cases to minimize the prediction error.

$$E^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - a - bx_i)^2 \tag{2.2}$$

Figure 4 below represents what a cross plot of an attribute and a target log may look like, including a computed regression line. It is easy to see that the prediction error is quite high here and would lead to poor prediction of the target log property. However, the use of multiple attributes may lead to the formation of a more robust statistical relationship between the attributes and the target log property.

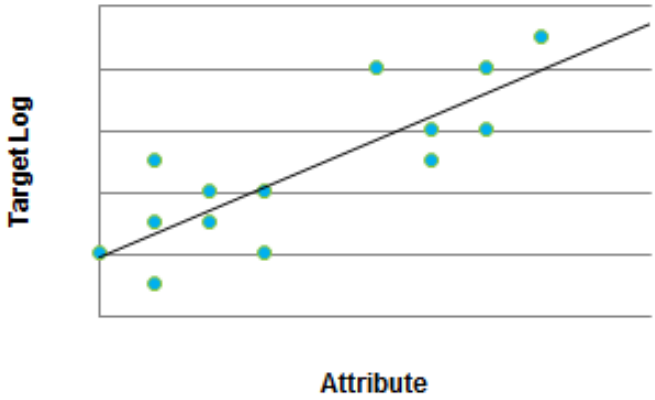


Figure 4: Hypothetical Cross-plot of a Single Attribute vs a Particular Target Log

Multiple attributes can be used to decrease the prediction error for a certain target log property. By extending equation 2.2, each target log sample is to be modeled as a linear combination of attribute samples at a particular time (Hampson, Schuelke, & Quirein, 2001). The log property is modeled with the linear equation 2.3. The weights in equation 2.3 can be solved for in the same manner as in equation 2.2, shown below in equation 2.4. The weights (w_o–$w\_$…) are solved using multiple linear regressions.

$$L(t) = w_o + w_1 A_1(t) + w_2 A_2(t) + w_3 A_3(t) + \cdots \tag{2.3}$$

$$E^2 = \frac{1}{N}\sum_{i=1}^{N}\left(L_i - w_o - w_1 A_1(t) - w_2 A_2(t) - w_3 A_3(t)\right)^2 \tag{2.4}$$

Equation 2.4 would be useful if every attribute had the same resolution as the target log property. However, every attribute may have a different resolution to one another as well as a lower sample rate compared to the well log property. Equation 2.5 is equivalent to

equation 2.4, but here the convolutional operator is used to reconcile the effect of the differing resolutions. Figure 5 illustrates this issue of resolution between the subset of attributes and the log property.

$$E^2 = \frac{1}{N}\sum_{i=1}^{N}(L_i - w_o - w_1 * A_{1i} - w_2 * A_{2i} - w_3 * A_{3i})^2 \qquad (2.5)$$

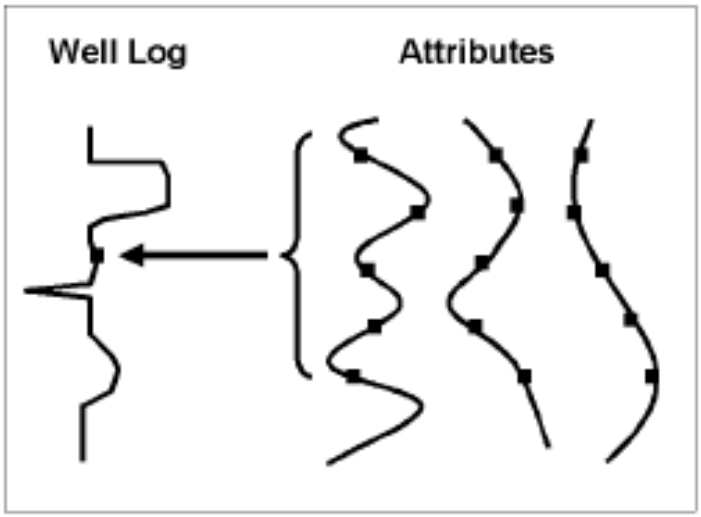


Figure 5: Well to Attribute Comparison (Hampson, Schuelke, & Quirein, 2001)

2.2 Multi Attribute Prediction of Rock Properties with Seismic Attributes

Deriving the optimal subset of attributes to predict a target log with is (theoretically) simple - use an exhaustive search of every combination of attributes to determine the set which results in the lowest prediction error. This, however, is computationally inefficient, especially if there are multiple volumetric attributes to predict with. Step-Wise Regression can be used to determine the attribute subset more efficiently. The process of step-wise regression is as follows:

Step 1: Find the best attribute by exhaustive search. The best attribute is chosen by virtue of it resulting in the lowest prediction error.

Step 2: Once the best attribute is found, find the best attribute of the remaining attributes by exhaustive search that pairs with it. The best pair of attributes will be chosen by having the lowest prediction error.

Step 3: Using the best attribute pair found in step 2, solve for the best triplet of attributes, etc.

The above process can be continued in order to find a larger subset of attributes. However, the process above does not guarantee that the best subset of attributes has been found, which would be the case if an exhaustive search had been used. The process of *Cross*

Validation will be used to determine how many attributes to use in the prediction. The subset of attributes from stepwise regression is as large as one wants. “Yet with too many attributes one can over train the prediction.” (Kalkomey, 1997)

A Multi-Attribute transform with N+1 attributes will always have a prediction error less than or equal to the transform with N attributes (Hampson, Schuelke, & Quirein, 2001). Cross validation is used to find “useless” attributes or those that are “over-training” the prediction. Cross Validation divides the training dataset into subsets:

Training Dataset: Used to derive the weights of the Transform (All Wells -1)

Validation Dataset: Used to Measure the Final Prediction Error (Hidden Well)

The validation dataset consists of the hidden well while the training dataset consists of all the other wells. This process is repeated for all wells. The total validation error is the root mean square average of the individual errors, represented in equation 2.6.

$$E_V{}^2 = \frac{1}{N}\sum_{i=1}^{1} e_{Vi}{}^2 \qquad (2.6)$$

The validation error for any number of attributes will always be higher than the prediction error because there are fewer wells involved in the training. Figure 6 shows an example of what an attribute validation error and prediction error may look like as a function of number

of attributes. One can interpret that any more than 3 attributes could be worse for this particular prediction.

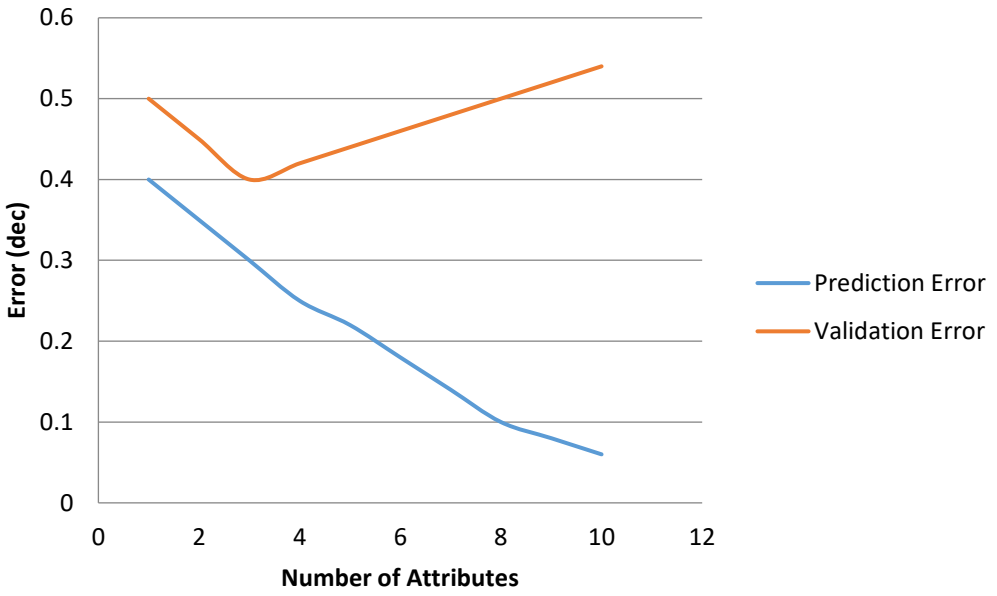


Figure 6: Cross-plot of the number Attributes vs. Error

Cross well validation is also used to determine if there are wells in the dataset that contain spurious data that the prediction would not benefit from. This needs to be examined in fine detail to ensure that only spurious data is excluded from the prediction and that which represents true geology is retained. An example of this analysis is shown above in Figure 7.

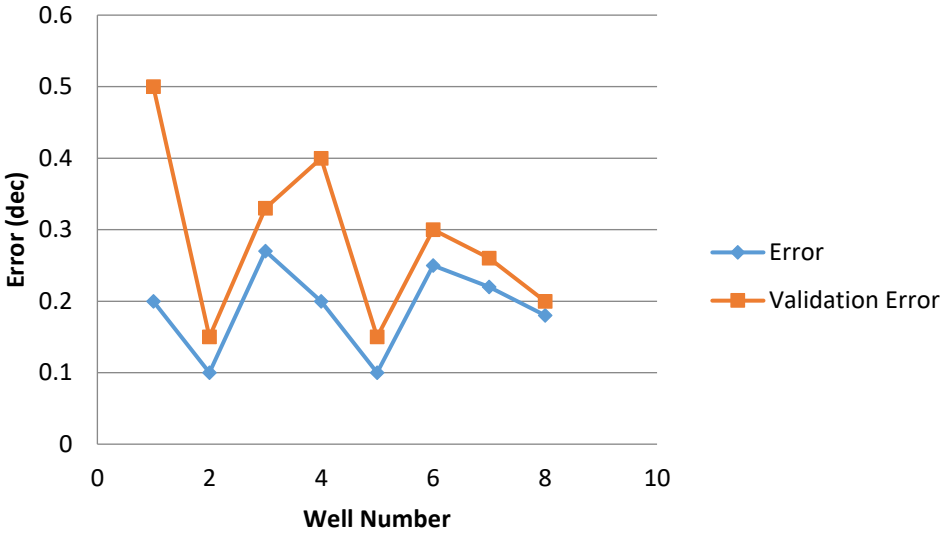


Figure 7: Cross-plot of each Well vs. Error

Chapter 3 Geostatistics

3.1 An Introduction to Geostatistics

The estimation of rock and reservoir properties has developed into an essential process in hydrocarbon exploration. In hydrocarbon exploration well data is largely considered "ground truth". Even before the invention of seismic there were processes to interpolate or estimate properties between or around wells. 3D seismic has the ability to "fill in" missing information between and around wells, albeit at a lower resolution.

Today one can divide the methods used to predict reservoir properties into two main classes. "Conventional methods for estimating rock and reservoir properties from

seismic data rely on empirical or regression formulas" (Tordorov, 2000). Geostatistics uses the ability to analyze and predict values associated with spatial relationships.

Both classes have the ability to use non-linear and linear regression, yet geostatistics relies on the ability to use spatial correlation to predict for the target property. There are positives and negatives of both classes. Using only Non-Linear regression described in the last chapter, there is no variation in weights, no measurement of spatial correlation. This creates areas of erratic change around well data, which might not be the case. Using a process such as cokriging to build the reservoir model is appealing to most geologists because the result is more realistic (e.g., channels look like channels). "However object-based methods have their limitations in that the algorithms require numerous input parameters describing the geometrical dimensions, such that object modeling becomes nearly deterministic, especially when many wells are used as conditioning information" (Chambers & Yarus, 2002). Geostatistical techniques have an "ability to provide improved reservoir description, considering the spatial correlation of the geophysical data, and add the ability to assess the uncertainty in the estimation process" (Tordorov, 2000).

3.2 Introduction to Statistical Concepts

Before understanding the more advanced geostatistical methods we will cover the basic statistical terms. "A discrete random variable, *Z*, is a variable that can take a series of outcomes (realizations) z_i , where $i=1,.....,N$, with a given set of probability of

occurrence p_i , where *I=1,…,N*" (Tordorov, 2000). The probability of the occurrence of N must satisfy these conditions shown in equation 3.1:

$$p_i \leq 1, for\ all\ i = 1, \ldots., N$$

$$\sum_{i=1}^{N} p_i = 1 \tag{3.1}$$

The probability – weighted sum of all possible occurrences is otherwise known as the expected value (Z), or the mean m shown in equation 3.2. "The variance, σ^2, of a random variable *Z* is defined as the expected squared deviation of *Z* about its mean" (Tordorov, 2000).

$$E(Z) = m = \sum_{i=1}^{N} p_i z_i \tag{3.2}$$

$$Var(Z) = \sigma^2 = E(Z - m) = \sum_{i=1}^{N} p_i (z_i - m)^2 \tag{3.3}$$

The variance is a measure of spread of a distribution around the mean. It is also important to note that the variance is not measured in the same units as the input variable, whereas the standard deviation is. "The covariance, *Cov(X,Y)*, of two random variables, *X* and *Y*, is defined as

$$Cov(X, Y) = \sigma_{XY} = E\left((X - m_x)(Y - m_y)\right) = E(XY) - m_x m_y \tag{3.4}$$

Where m_x and m_y are the mean values of the two random variables" (Tordorov, 2000). Now the dimensionless version of the covariance is otherwise known as the correlation coefficient.

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(x)Var(Y)}} \tag{3.5}$$

"Correlation is a measure of linear dependence between two random variables and is confined in the interval [-1,1]" (Tordorov, 2000). In geostatistics the method by which we measure variability of two random variables, *X* and *Y*, is to use a Variogram. A semivariogram is one type of variogram. Semivariance is a measure of the dissimilarity (Bohling, 2005). The greater the value of dissimilarity the less related the two variables are.

$$\gamma_{XY} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \tag{3.6}$$

3.3 Evaluating Spatial Variability

Most geostatistical techniques rely on the ability to produce a way of representing spatial continuity in the dataset. Thus there is an assumption that when viewing a certain map often low values tend to be near other low values and high values tend to be near other high values. This means spatial continuity exists in the map. Thus the assumption can be made that two data points close to each other have the probability of being more similar

than two points with greater distance between them. “The dissimilarity between all the data pairs can be expressed by the variogram as a function of the distance *h*, often called offset or lag, between them. Where *N(h)* is the number of all pairs separated by distance *h*” (Tordorov, 2000).

$$\gamma^*(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} \left(Z(x_i) - Z(x_i + h)\right)^2 \tag{3.7}$$

In production it is common to put the data points into different offset bins and calculate the variogram value in each bin. This would make *h*, the lag, define the offset to the center of the bin. Lag can be seen visually in figure 8 below. There are several other terms below that characterize the experimental variogram.

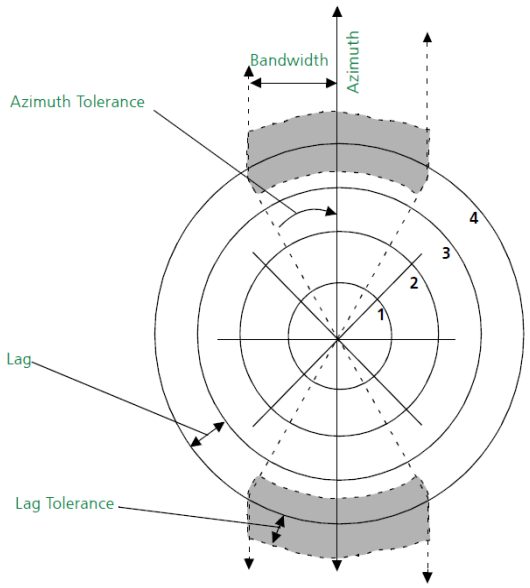


Figure 8: Attributes of a Variogram *(Paradigm, 2012)*

Take for example *Azimuth*. If the azimuth range increases for a single variogram then that would allow more data points into the variogram. Sometimes the best way is to model different azimuths with different variograms. This will be displayed later when dealing with the real data. Now once the experimental variograms are made the next step is to fit the best model to them. The basic schematics of a semivariogram are shown in figure 9 below. Common variogram modeling techniques include spherical, exponential, Gaussian and power.

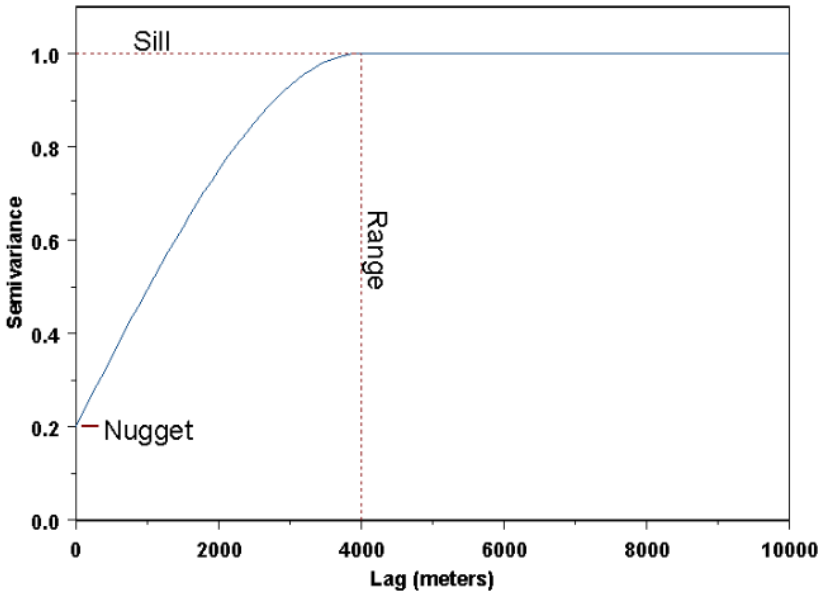


Figure 9: Schematics of a Semivariogram (Bohling, 2005)

Sill: The Cutoff for the variance in the model.

Range: The Actual Distance at which the variogram reaches the Sill.

Nugget: The nugget represents variability at distances smaller than the typical sample spacing, including measurement error.

An example of what a model would look like would be the exponential example above where C is the sill, and a is the range. “Now the covariance can be computed from the variogram where $\gamma(\infty)$ is the sill and C(0) is the zero-offset covariance” (Journel, 1989).

$$\gamma(h) = C\left[1 - \exp(-{}^{h}/_{a})\right] \tag{3.8}$$

$$Cov(h) = \gamma(\infty) - \gamma(\mathrm{h}) = \mathrm{Cov}(0) - \gamma(\mathrm{h}) \tag{3.9}$$

3.4 Kriging

Kriging is used throughout the industry as a basic statistical interpolation technique. The application of kriging has a single purpose, to estimate a property at an unmeasured location using a linearly weighted sum of the known values. Equation 3.10 represents the basic equation used in the simple kriging technique. There are multiple kriging techniques, such as Simple, Ordinary, Indicator and Universal kriging. Since the main geostatistical tool we will be using in this report is cokriging, there will only be a brief explanation of simple kriging. When z_0 is the property to predict for, the equation below describes the linear estimator for N known values. The estimation error at any given location is the difference between the estimated value, z_0*, and true value, z_0.

$$z_0{}^{*} = \sum_{i=1}^{N} w_i z_i \tag{3.10}$$

$z_0{}^{*} = Estimate\ at\ the\ Grid\ Node\ Location$

$w_i = The\ unknown\ weight\ assigned\ to\ the\ sample\ z_i$

$z_i = Regionalized\ Variable\ at\ a\ given\ location$

The bias is the mean of the error distribution. An estimation technique typically seeks to produce unbiased results, which would require in this case that the mean of the error distribution (or bias) be zero.

$$\frac{1}{K}\sum_{i=1}^{K}(z_{0i}{}^{*} - z_{0i}) = 0 \tag{3.11}$$

However, we do not know the true values. The way around this is to define the unknown values as an outcome of a random variable. "For any point at which we attempt to estimate the unknown value, our model is a random function that consists of random variables, one for each of the sampled locations, Z_i , i=1….N, and one for the point we are trying to estimate, $Z_0{}^{*}$

$$Z_0{}^{*} = \sum_{i=1}^{N} w_i Z_i \tag{3.12}$$

Now it is possible to evaluate the estimation error. The estimation error is the difference between the predicted Z^{*} and the random variable Z which is the model of the true value. Stated before, the estimator is unbiased if the expected value of the estimation error is zero.

$$E(Z_0{}^{*} - Z_0) = 0 \tag{3.13}$$

The unbiased estimator can be achieved only if the weight is not 0, and the sum of all weights equals one. The kriging estimator is assumed to be the best because the weights are determined by the "minimization of the error variance" (Tordorov, 2000).

$$Var(Z_0{}^{*} - Z_0) \rightarrow min$$

"The set of weights that minimize the error variance under the constraint that they sum to one satisfies the following N equations" (Isaaks & Srivastava, 1989).

$$\sum_{j=1}^{N} w_j Cov\{Z_i, Z_j\} + \mu = Cov\{Z_0, Z_i\} \qquad i = 1, \ldots\ldots, N \qquad (3.14)$$

Where μ is the Lagrange operator and $Cov\{\}$ is the covariance function. The modeled variogram constructs the covariance values used. The ordinary kriging estimate is created from the weights solved with the above equation.

3.5 Cokriging

The kriging methods described in the previous section are not multivariate models in the basic statistical sense. They do however employ random function models which are made up of infinite numbers of random variables. Yet they are creating these random variables from a model of one attribute. If a secondary dataset that has mutual spatial behavior with the log data is available then a process such as cokriging is possible.

"The main purpose of cokriging is to work as a true multivariate estimation method able to deal simultaneously with two or more attributes defined over the same domain, which is called a coregionalization" (Olea, 2009). The ordinary cokriging estimator Z_0 is written as

$$z_0{}^* = \sum_{i=1}^{N} w_i z_i + \sum_{j=1}^{N} \beta_j t_j \qquad (3.15)$$

$z_0{}^* = Estimate\ at\ the\ Grid\ Node\ Location$

$w_i = The\ unknown\ weight\ assigned\ to\ the\ sample\ z_i$

$z_i = Regionalized\ Variable\ at\ a\ given\ location$

$t_j = The\ known\ secondary\ data\ that\ is\ co-located\ with\ the\ z_i$

$\beta_j = The\ unknown\ weight\ assigned\ to\ the\ sample\ t_j$

The estimate requires the spatial covariance model of the hard data, the spatial covariance model of the soft data, and the spatial cross-covariance model of both the hard and soft data. For every additional soft data attribute there must be an additional covariance model. Collocated cokriging has a benefit over cokriging which stems from the assumption that the soft data is not exactly coincident with the hard data. Thus, a reduced form of the collocated cokriging formula that keeps the covariance of the secondary data at the target node location:

$$z_0{}^* = \sum_{i=1}^{N} w_i z_i + \beta_0 t_0 \tag{3.16}$$

Cokriging is similar to the nonlinear regression process stated in the previous chapter such that multiple volumetric attributes can be used to predict a particular target log. There are multiple differences between the two techniques used in this research. The main difference is that cokriging does not exactly assume spatial continuity. A variogram is used to measure where the hard data is trusted more than the soft data. This gives cokriging the ability to have different sets of weights for different grid node positions. In the non-linear regression process the weights are the same throughout the entire volume, assuming spatial continuity. This may seem like a positive for using the cokriging method

over the non-linear regression method, but this is not exactly the case. A variogram must be built for every attribute, every attribute compared to the next, and the hard data compared to every attribute. That is a lot of modeling for a single prediction. In the end the modeling of every variogram may lead to a result that is rather “cherry picked”. The point of the research however is not to compare the non-linear regression process and the cokriging process but to understand how to calculate uncertainty in these processes. The nonlinear-regression process has the ability to calculate error from cross validation techniques, blind well testing. However it does not have the ability to calculate spatial uncertainty. Cokriging does have the ability to calculate spatial uncertainty. Thus the nonlinear regression will produce the target log output volume. Then this target log volume will be used as the soft data for the cokriging application. The collocated cokriging process will be used alongside sequential Gaussian simulation to produce the best result for uncertainty measurement.

3.6 Sequential Gaussian Simulation

In geostatistics, simulations are used to characterize heterogeneous reservoirs. The benefit of using a simulation over a traditional interpolation method is that the simulation is a stochastic approach. This leads to the ability to calculate many realizations, or solutions, to a problem. Simulations add a certain random residual to the kriged value at any given point to simulate the uncertainty of that value (Paradigm, 2012). The result of the simulation usually does not have the smooth output that ordinary kriging or cokriging produces. A common place geostatistical simulation technique is Sequential Gaussian

simulation. The term sequential indicates the simulated values will be used as the input for the simulation a later point. Gaussian describes the probability distribution the simulation is sampling from. Generally multiple simulations are run to gather a robust statistical result. With these many solutions one has the ability to identify reliability and uncertainty in the geostatistical method.

Steps in a Sequential Gaussian Simulation (Tordorov, 2000)

1) Define a Grid and Insert the Known Data at their Locations
2) Select a random point and estimate its value using kriging or cokriging
3) The estimate consists of a value and a variance; assuming a Gaussian distribution around the estimated value, determine a new value for the point by use of a random number generator
4) Repeat the above steps for every point, treating the previously estimated ones as exact, i.e. use them in the kriging (cokriging) estimation

Once the target number of realizations has been computed an uncertainty can be computed. First the mean will be calculated from all of the realizations.

$$m = \frac{1}{n}\sum_{i=1}^{n} z_i \tag{3.17}$$

The mean will be much smoother than each realization separately. Thus the mean may not represent the best realization to predict with. However with the mean calculated the variance and the standard deviation can be calculated.

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - m)^2 \tag{3.18}$$

Chapter 4 Research Dataset

4.1 Introduction to Data

The Dataset that will be used in this research will be the Buffalohorn ResSCAN Dataset, courtesy of Ion Geophysical. The Dataset was acquired in north central Oklahoma. Figure 1 in the project overview shows the relative area of acquisition. Table 1 below illustrates some of the key acquisition information. The seismic dataset was acquired with multi-component receivers. The data was processed with P and C wave components. The bulk of the work in producing uncertainty estimation will be conducted using the P wave component data. However the C wave data will be used in the final product when using real data.

Table 1: Acquisition Parameters for the 3D Seismic, otherwise known as the BuffaloHorn Scan

Recording Parameters	
Record Length	6 Seconds
Sample Rate	2 Milliseconds
Source Parameters	
Primary Source Type	Vibroseis - 64,000 lbs.
Number of vibrators per fleet	3
Design Attributes	
Nominal Fold	338
Bin Size	82.5 ft x 82.5 ft
Minimum Offset	117 ft
Maximum Offset	12,017 ft

There are a large number of wells drilled in the area. Due to proprietary reasons, this will be limited to 17 used in this research. Each well contains a full suite of log information that has already been processed, and will be assumed to be correct. All 17 wells are located inside of the survey and sparsely distributed across the survey. Figure 2 in the project overview shows the spacing of the wells within the boundary of the seismic data.

There are multiple areas of interest in this survey area. The Pennsylvanian sandstone is the common conventional rock layer in the area. The other units of interest are mostly unconventional in petroleum exploration. The main rock units of focus in this research will be the Mississippi Limestone and Woodford Shale units.

4.2 Geologic Information

The area of interest for this research is the upper middle Paleozoic. The Periods include the Lower Pennsylvanian, Mississippian and Upper Devonian. In the region the middle to late Devonian was comprised of "a period of widespread uplift and erosion" (Johnson, 2008). This was the foundation of the Woodford shale. The Woodford shale is a high TOC marine shale. "Shallow seas covered most of Oklahoma during most of the first half of the Mississippian Period" (Johnson, 2008). The Mississippian is made up of mostly limestones and cherts in the area. Above the Mississippian is the Pennsylvanian Sandstone unit. This boundary marks a big unconformity. Structurally the area of research is just east of the Nemaha Ridge. In the area are significant unconformities between the top and base of the

Mississippi Limestone. Within the seismic section there is a northeast-trending, normal fault zone (Wickstrom & Johnson, 2012). This zone can be seen in the coherence slice shown below in figure 10.

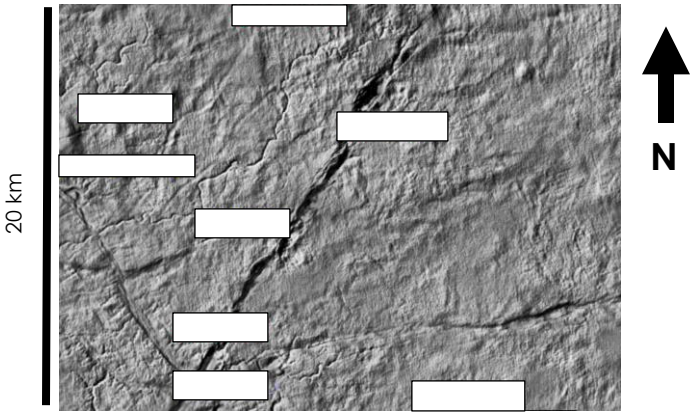


Figure 10: Coherence Time Slice of Seismic Volume, white boxes: hidden well names

(Greyscale Image, Darker area is area of lower coherence)

The Mississippian Limestone is broken up into three layers in the survey. At the top of the section is a low resistivity, very high porosity limestone. This unit is very rare in the seismic section. The middle unit is comprised of a diagenetic chert, which has moderate resistivity and porosity. This section is very common in the seismic section. The bottom of the section consists of an unaltered limestone with variable resistivity, low porosity, and is common in the seismic section. Figure 11 illustrates what an example log would look like for this seismic section.

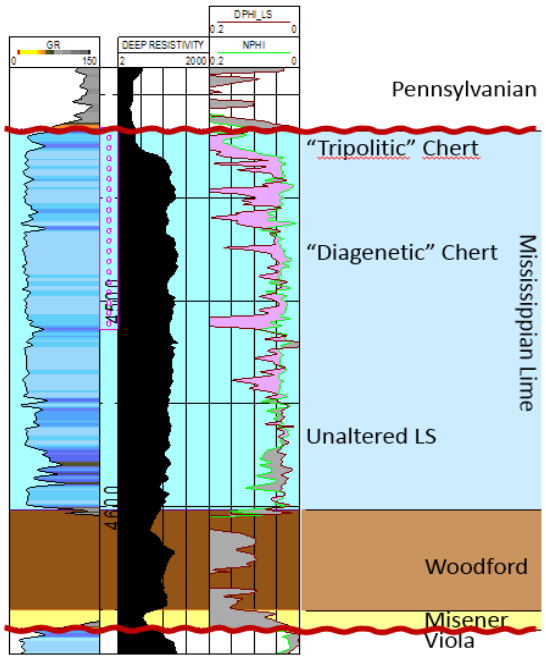


Figure 11: Type Log in Area of Interest

4.3 Creation of Synthetic Dataset

Verifying a process that quantifies the uncertainty or error standard deviation of a prediction is impossible to complete without knowing the answer first. A synthetic dataset is to be created to verify the process of quantifying uncertainty in the property estimation. When predicting for a reservoir property with the synthetic dataset the answer is already known, so production of a related uncertainty volume is straight forward.

To stay relative to the geology of the real dataset the synthetic dataset was constructed from the PP-PS inversion of the real 3D seismic data. The inversion outputs used to create the 3D synthetic seismic volume include acoustic impedance, shear impedance, Vp/Vs and density. Each one of these attributes has been scaled to pp-time.

Below is a picture of the real data compared to the synthetic data. As one might expect the real data has a lot more definition than the synthetic.

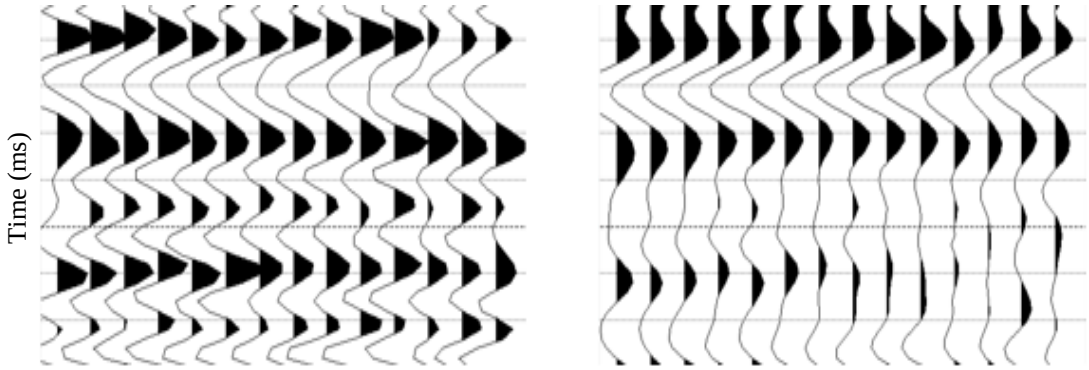


Figure 12: Real vs. Synthetic Data, Respectively in Time(ms)

The next step was to make the logs to predict for in the multi-nonlinear regression and since we want the answer to be known we selected one of the attributes that was used to create the synthetic dataset. The attribute that was chosen was the Vp/Vs ratio of the dataset. So a composite trace was extracted from the Vp/Vs volume at each well location, corresponding to the real well locations.

Chapter 5 Synthetic Dataset Application

5.1 Prediction of Log Property using Non-Linear Regression

The point of using the synthetic dataset in the process is to confirm how to measure uncertainty when the answer is known. Thus the log to predict for must be a known quantity. The inputs that were used to create the synthetic dataset are acoustic impedance, shear impedance, Vp/Vs and density. The chosen log property to predict for is Vp/Vs ratio. This log property was chosen so that one of the external attributes to the inversion could

be either Vp or Vs, but only one and not the other. This would give the inversion a good prediction, but not a perfect prediction.

Figure 13 is an example of what the Vp/Vs log would look like for a single well location compared to the seismic trace, density log, acoustic impedance, shear impedance and different angles of the seismic trace. From here one may notice similarities between the blue traces (attributes) and the red trace (log to predict). The next step is to figure out the single best attribute that predicts for the Vp/Vs log.

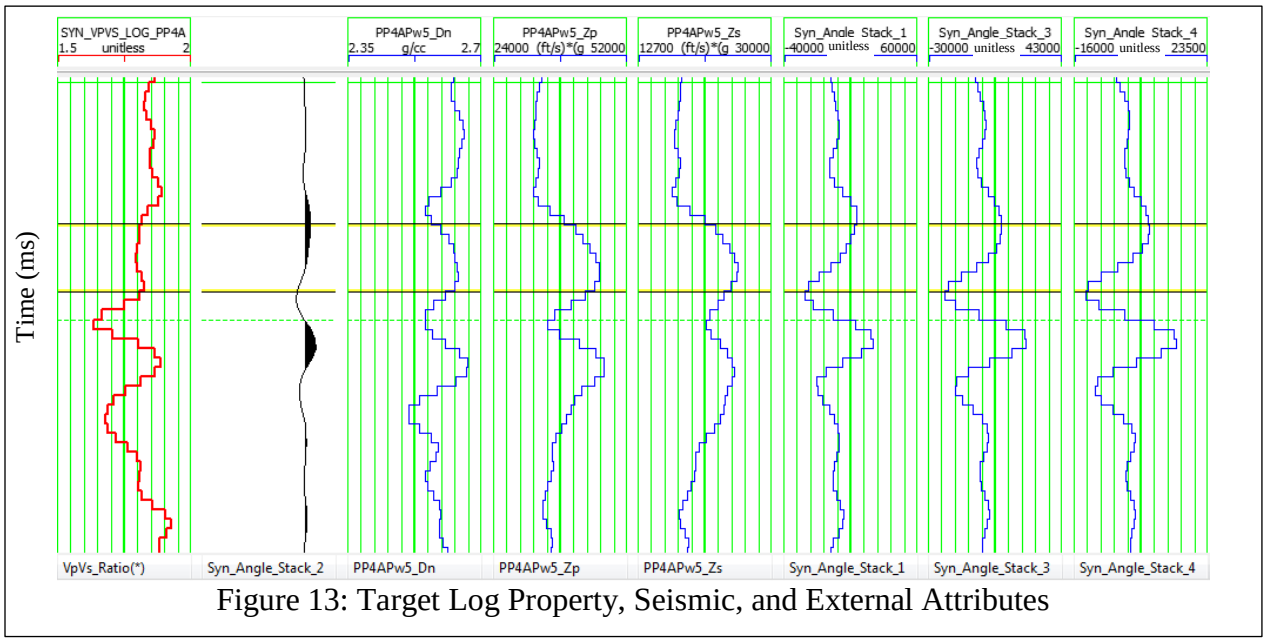


Figure 13: Target Log Property, Seismic, and External Attributes

Table 2 below contains a list of attributes that were compared to the Vp/Vs Ratio. As one may notice the attribute could be directly compared to the Vp/Vs ratio (like number 1 in the list). Or the target may have a nonlinear transform applied to it and the attribute also has a nonlinear transform applied (as number 3 in the list). The best attribute that

predicts for the Vp/Vs ratio at all of the well log locations was found to be the square root of the Shear Impedance (this was the attribute with the lowest prediction error).

Table 2: Attribute List of the Ranked as the single best attribute is number 1, the lower the rank the higher the training error of the particular attribute

Rank	Target	Attribute	Training Error	Correlation
1	V_P/V_S	$\sqrt{SI}$	0.034003	0.846217
2	$\left(V_P/V_S\right)^2$	$\sqrt{SI}$	0.035494	0.824654
3	$\left(V_P/V_S\right)^2$	$Density$	0.035602	0.824023
4	$\sqrt{V_P/V_S}$	AI	0.038602	0.809451
5	$\log\left(V_P/V_S\right)$	SI	0.038941	0.801239
6	$1/\left(V_P/V_S\right)$	Bandpass Filter (25,30-35,40)	0.039742	0.794981

From table 2 we can infer that the square root of shear impedance is the best attribute to predict with. Figure 13 below shows what a prediction would look like for the Vp/Vs ratio using this single attribute. The black log is the ground truth and the red log is the prediction. The final prediction was only made in the area of interest, which in this case is the Mississippi Limestone. This area is labeled by the top and bottom horizontal lines on each well track. The figure only shows the results for three wells, but this prediction was made for all seventeen wells. At this point different shifts, smoothing techniques and added

processing can be performed on the target logs to provide a better possible prediction. However for the synthetic dataset there will be no added processing.

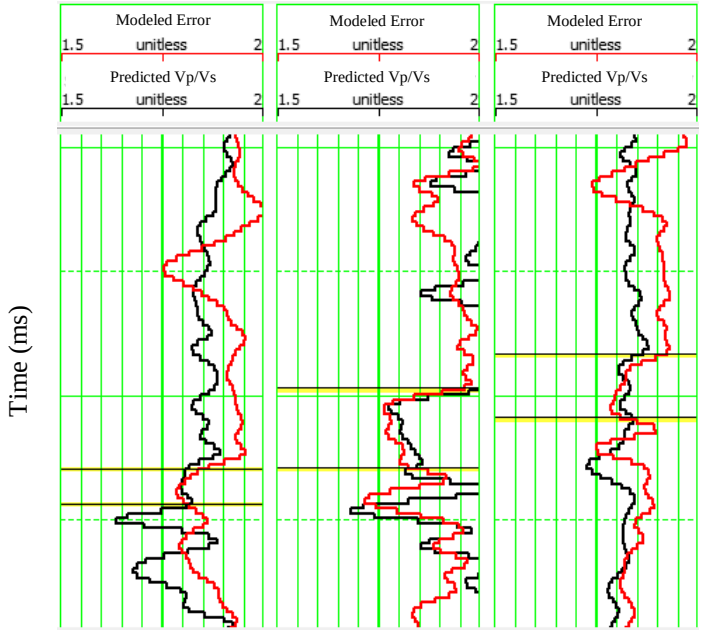


Figure 14: Prediction of Vp/Vs using Best Single Attribute

Now to see how multiple attributes can help predict for this target log, Vp/Vs. To figure out the best subset of attributes, stepwise regression will be used. Instead of picking the best subset based on prediction error, validation error will be used to decide what the best subset of attributes is. This process will take a bit longer computationally but will provide the best answer in the end. To keep it simple a maximum of three attributes will be chosen to complete the subset. It is good to note here that even though the list in table 2

shows the best single attributes, this does not mean the top three are the best subset. Table 3 is a list of the three best attributes.

Table 3: Attribute List of the subset of attributes going into the multi-attribute prediction for the target log

Rank	Target	Attribute	Training Error	Validation Error
1	V_P/V_S	$\sqrt{SI}$	0.034003	0.036107
2	V_P/V_S	$Full\ Angle\ Stack$ {25,30 | 35,40}	0.030636	0.032796
3	V_P/V_S	Full Angle Stack Dominant Frequency	0.028639	0.031803

The aforementioned statement of using too many attributes to predict for a target log will be verified here. Adding attributes to a subset to predict with will always decrease the prediction error. However at some point when adding additional attributes the validation error can increase. Remember that the validation error is created using blind well testing. Figure 15 below shows the prediction error in black, and the validation error in red. As one can see the validation error does not increase with additional attributes. The slope of the validation lines does seem to decrease, meaning that the validation error will most probably increase with more attributes.

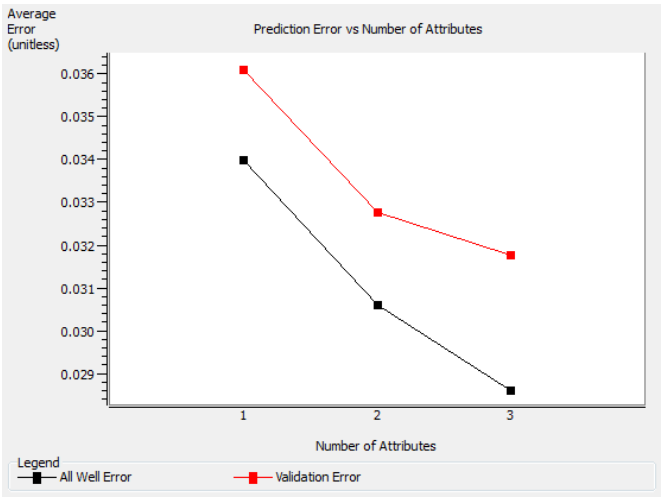


Figure 15: Prediction Error vs Number of Attributes

A relatively quick way of reviewing the prediction can be done using a cross plot. Figure 16 shows the cross plot between the target logs and the predicted logs. One may notice the prediction fits the trend line quite well. Each color on the cross plot represent a different well log and sample point.

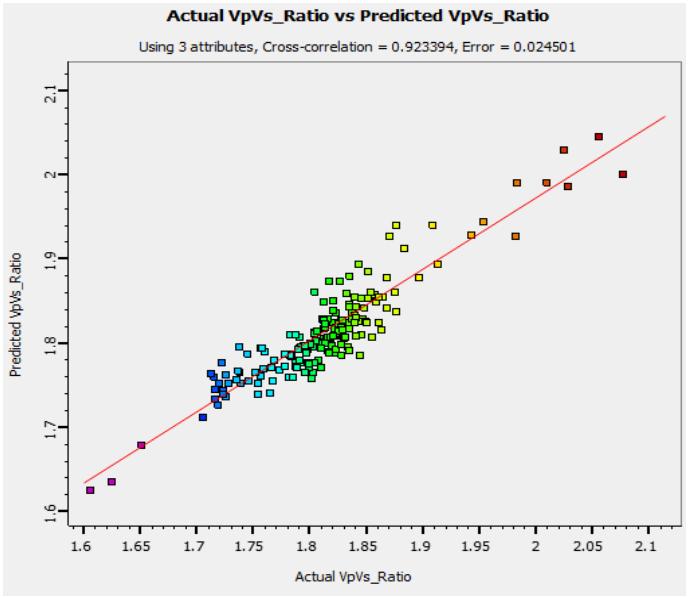


Figure 16: Actual vs Predicted Vp/Vs Ratio (unitless)

The prediction error of using a single attribute compared to using a subset of attributes has decreased by about one third. Figure 17 below shows a comparison between the multi-attribute prediction, on the right, and the single attribute prediction, on the left. The multi attribute prediction fits the wells in the more complex areas better. Yet the differences here are somewhat minor.

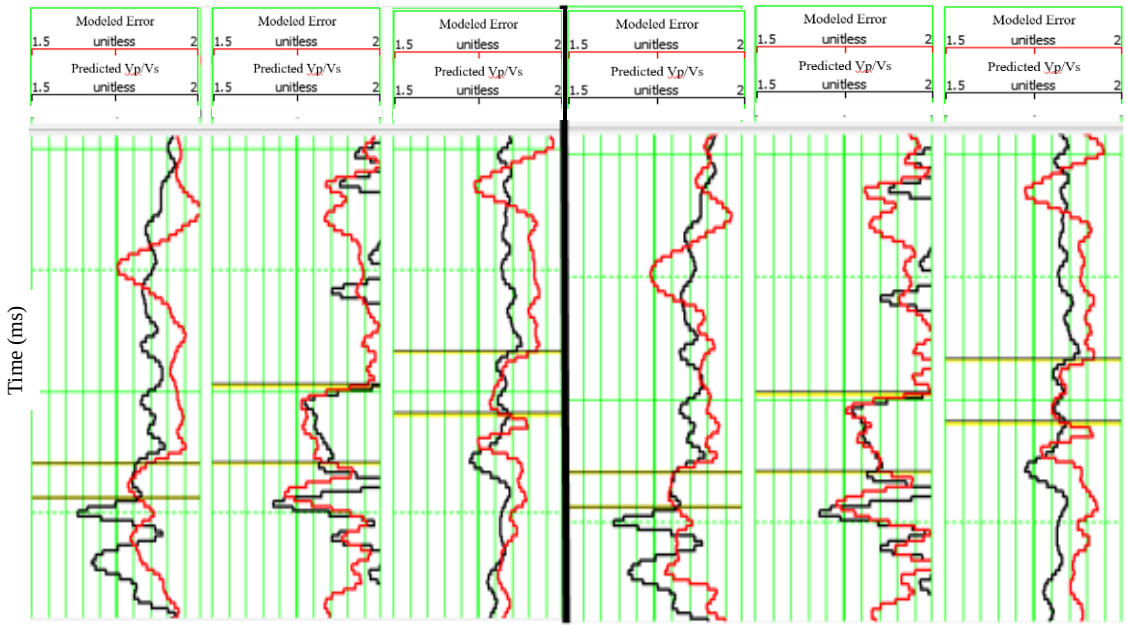


Figure 17: Single Attribute Prediction vs Multi-Attribute Prediction, Respectively

In the prediction process, cross validation has been used to accurately identify the best number of attributes to use. Cross validation will also be used to figure out what wells are predicting right. Figure 18 shows such a graph that represents the prediction and validation error for each well. One may notice some wells have a much higher error than others. This is where one would take out a well that had high validation error and see if the prediction is better without the well. However, in the case of the synthetic dataset we will

keep all the wells for the prediction. Remember also that the error is shown in units of measure for Vp/Vs. So a difference in error between 0.015 and 0.0425 is not an extreme amount.

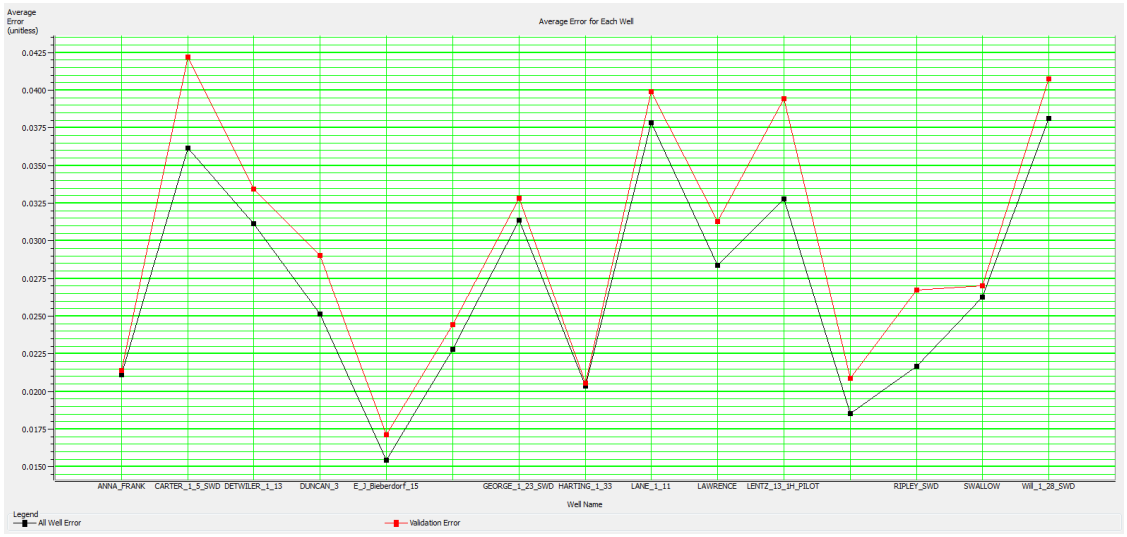


Figure 18: Average prediction error per well

The next step is to compare the predicted output and the control, or known output. Figure 19 and 20 are slices that follow along the Mississippi horizon. Figure 19 represents the predicted result and Figure 20 represents the control. As one would expect the prediction did not work everywhere. The western (left) side of the data is predicting much better than the eastern (right) side of the data. This may be due to the fact that there is more well data in a denser area in the west, whereas although the eastern section does have a lot of well data it is sparsely distributed. Figures 21 and 22 are cross section views between two different well logs. Figure 21 are the wells Will and Ernest, and Figure 22 are wells Harting and Carter. From the left to right one can see the initial seismic, the predicted vp/vs, and the control vp/vs. Between wells Will and Ernest the prediction is doing a very good

job. However in a more structurally complex area, predictions for Harting and Carter are not performing as well.

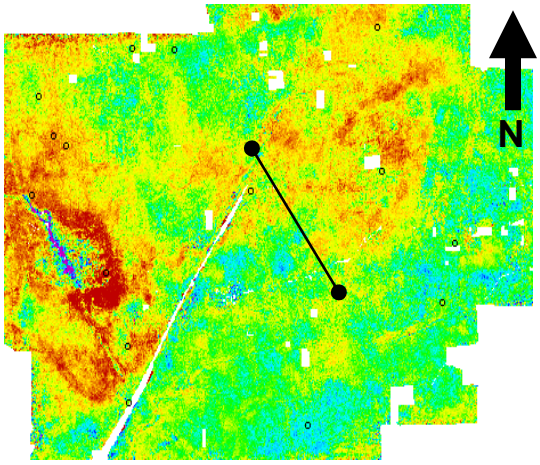


Figure 19: Non-Linear Multi-Variate Prediction of Vp/Vs, Lina A Represented

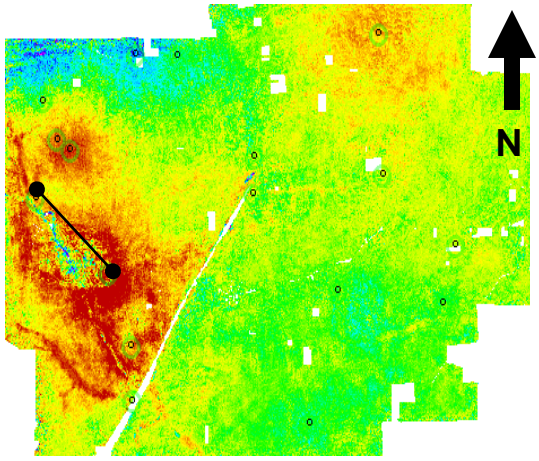


Figure 20: Control Vp/Vs, Line B Represented

50 km

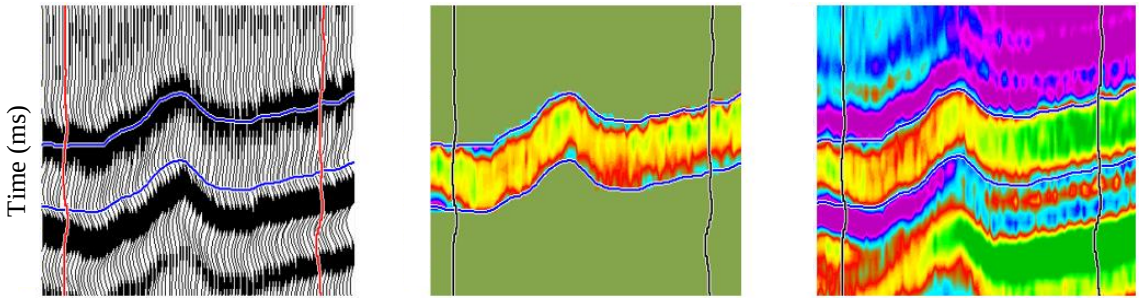


Figure 21: Seismic Stack, Predicted Vp/Vs, and Control Vp/Vs. (Line A : Figure 19)

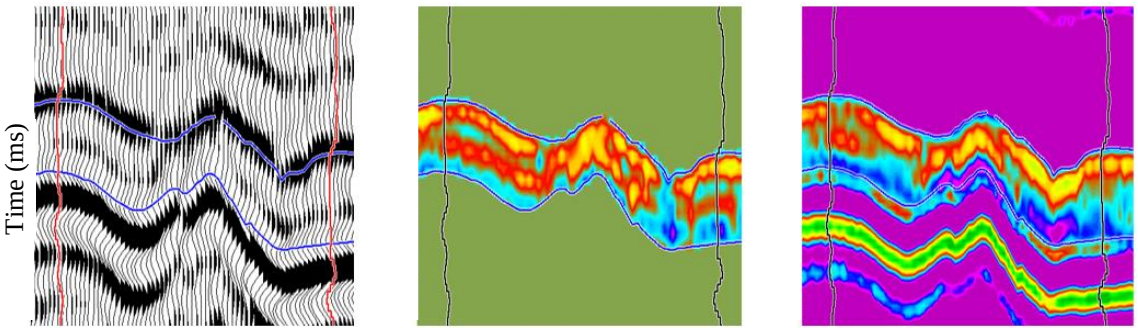
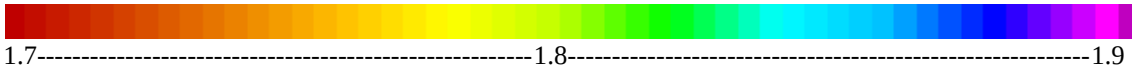


Figure 22: Seismic Stack, Predicted Vp/Vs, and Control Vp/Vs. (Line B : Figure 20)



Color Bar for Figures 19 through 22 (Vp/Vs, unitless)

5.2 Prediction of Log Property using Geostatistics

The Vp/Vs predicted volume from the non-linear regression follows similar trends to the Actual Vp/Vs. It is easy to validate where and how much error there is in the predicted volume due to the fact that the answer is known. However in the real dataset the answer will obviously not be known. The question is then how do you find the error in the volume away from the well without knowing the truth. First there needs to be an understanding of what is going on spatially in the data. How one well correlates to another spatially. Spatial correlation is not a necessity when predicting for a volume using non-linear regression. Yet the use of geostatistics can be applied to help understand how this data varies spatially in a quantitative fashion. Cokriging gives us the ability to predict for a property using spatial correlation and a covariate.

To start any geostatistical process some type of grid must be made. The grid should be sampled at least as fine as the sample rate of the hard data. In the case of the synthetic

dataset the hard data, well data, is at the same sample rate of the seismic data. So the grid will be made at the same resolution as the seismic data. The grid will be made up of 17,000,000 cells. After the grid is built the next step is to fill the grid with the non-linear regression predicted volume. Since the data is at the same sample rate as the grid there should be no reason to interpolate anywhere in the grid.

There are multiple inputs needed for collocated cokriging. The most important being the hard data, soft data, variogram and correlation coefficient. The only variable needed to compute is the variogram, since the correlation coefficient was calculated in the NLR.

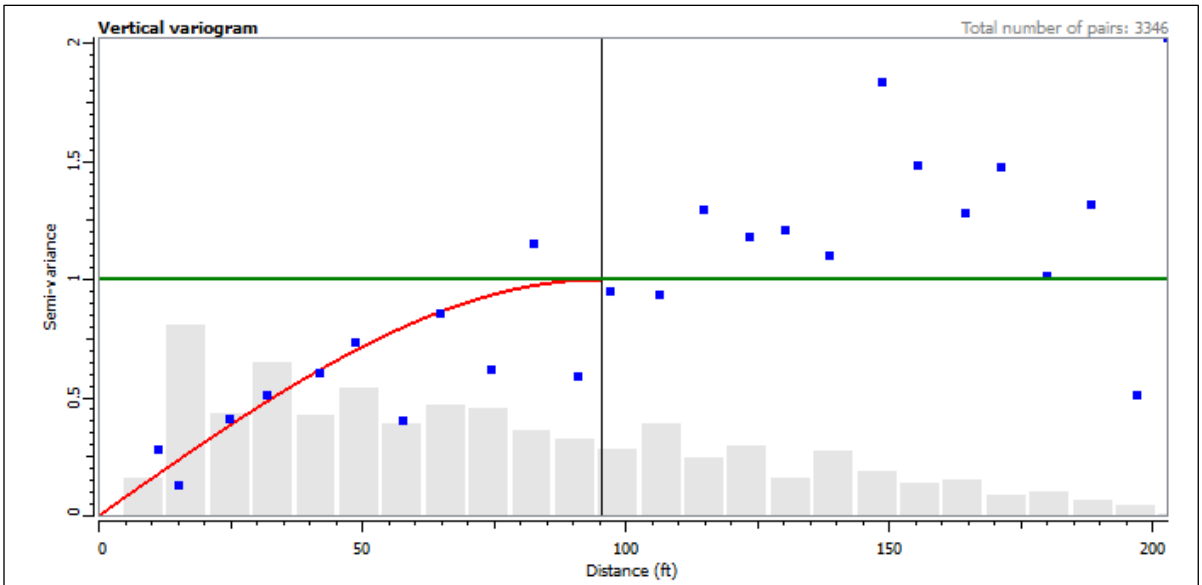


Figure 23: Vertical Variogram

The input variogram for collocated cokriging will be based on the hard data only. The methodology and terms for building a variogram model was illustrated in chapter 3. The collocated cokriging will be performed in 3D, so an understanding of spatial

correlation in three dimensions would be optimal. The three dimensions in the variogram would include vertical, azimuthal and dip. However for the synthetic dataset the variogram will be constructed only taking into account vertical and azimuthal correlation. This should not be too much of an issue considering the Mississippi limestone is quite flat in most areas. The variogram will only be built from the region of interest, the Mississippi Limestone.

Figure 23 shows the vertical variogram for the 17 wells with the chosen optimal model. The Model that best fits the hard data was a spherical model. The vertical variogram represents the data with variation only in distance. In the areal plane the model is also spherical with a major azimuth of 132 degrees. Figure 24 represents the azimuthal variogram showing two different azimuths. Figure 25 also displays the azimuthal variogram in an omnidirectional manner, which is a summary of all directions.

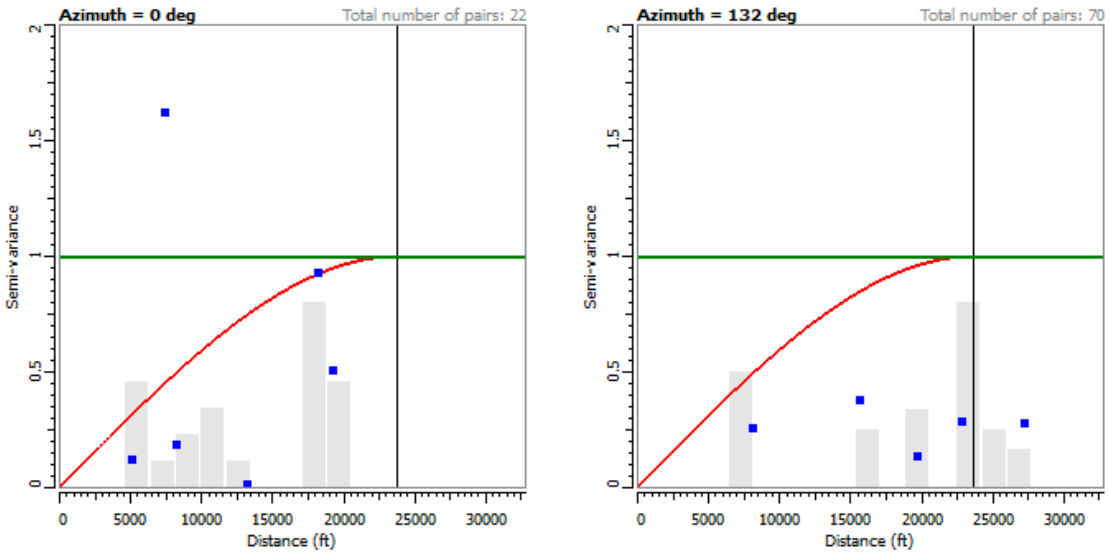


Figure 24: Variogram of Azimuth at 0 and 132 degrees, respectively

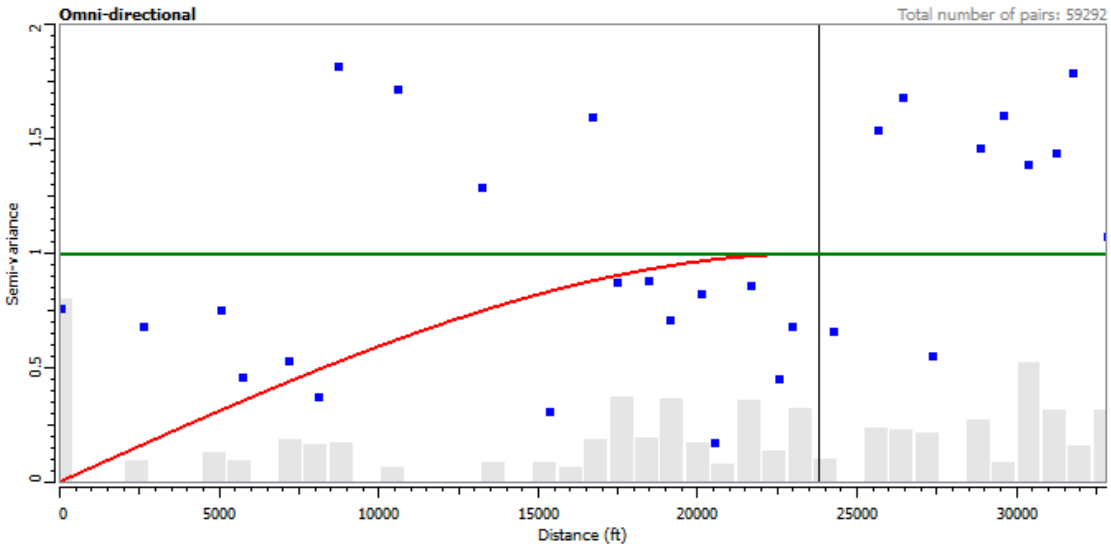


Figure 25: Omni-Directional Variogram

The next step is to define the histogram for the sequential Gaussian simulation (SGS). The histogram should represent the hard data for the cokriging result. Thus before we create a distribution for the simulation we should check the distribution of the hard and soft data. Figures 26 through 28 show the histograms of the Non-Linear Regression result, well logs and the control data, respectively. The *Emerge* output and the control data have a lot more data points than the well log data, so the histograms will look a lot smoother. The Emerge output and the control data have similar histograms, with the Emerge data trending more on the high side, and the control data trending more on the low side.

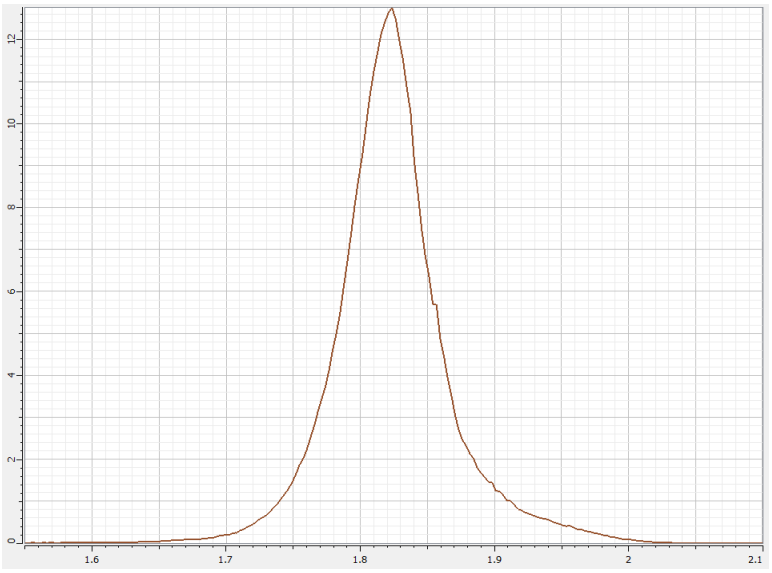


Figure 26: Histogram of Emerge output (Distance vs Frequency)

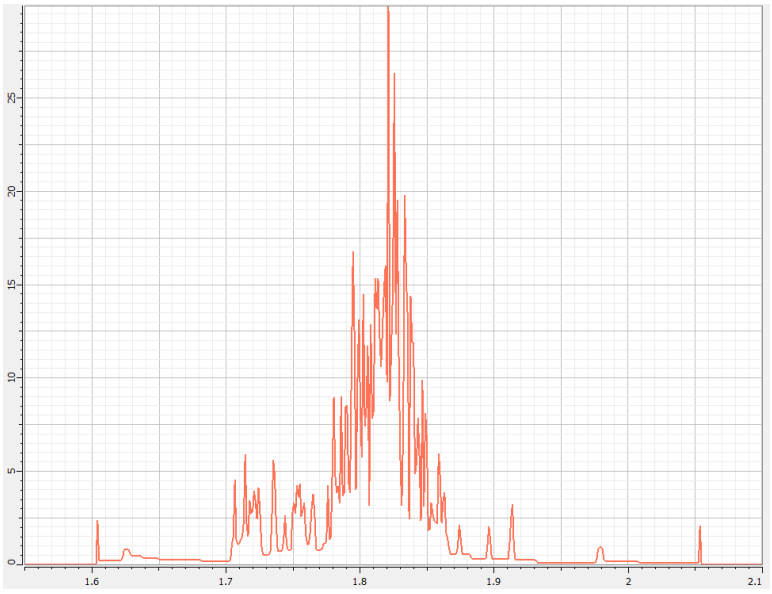


Figure 27: Histogram of Wells (Distance vs Frequency)

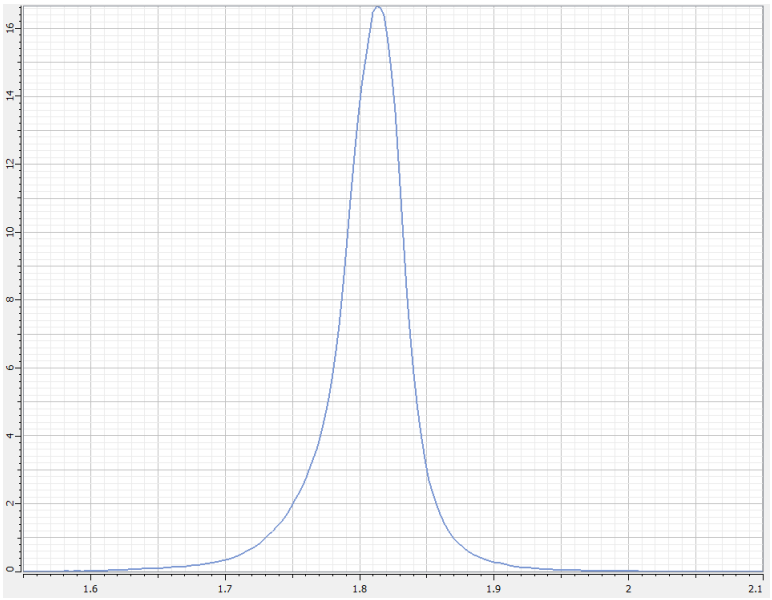


Figure 28: Histogram of Control Data (Distance vs Frequency)

The histogram that will be used for the input to the SGS will be a smoothed version of the well histogram. Now that all of the variables are accounted for we can start setting up the simulation. There will be 30 realizations run through the simulation. Figures 29 through 31 show a few of the output realizations. Notice that each realization is similar to the next with subtle differences. One may notice that most the variance from realization to realization is away from the well locations. This is due to the modeled variogram. The modeled variogram along with the correlation coefficient show how far away from the well the soft data should be trusted over the hard data.

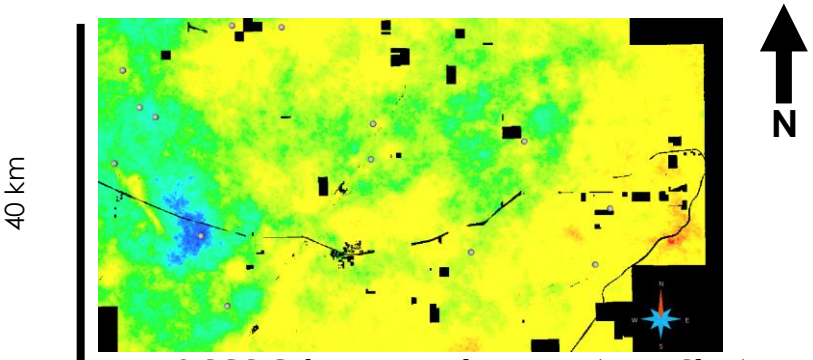


Figure 29:SGS Cokriging, Realization 1 (Time Slice)

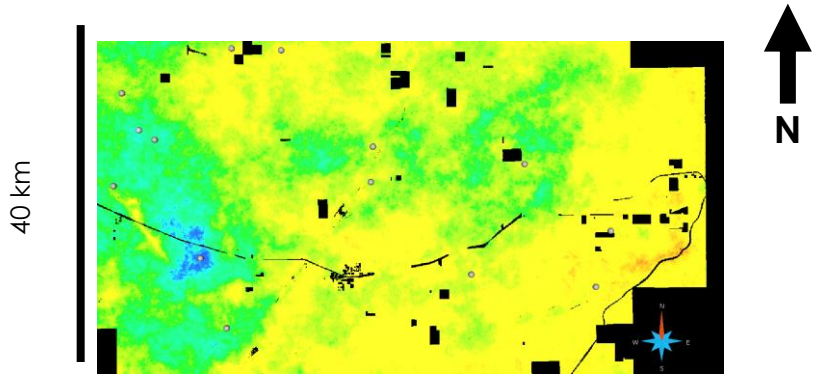


Figure 30: SGS Cokriging, Realization 15 (Time Slice)

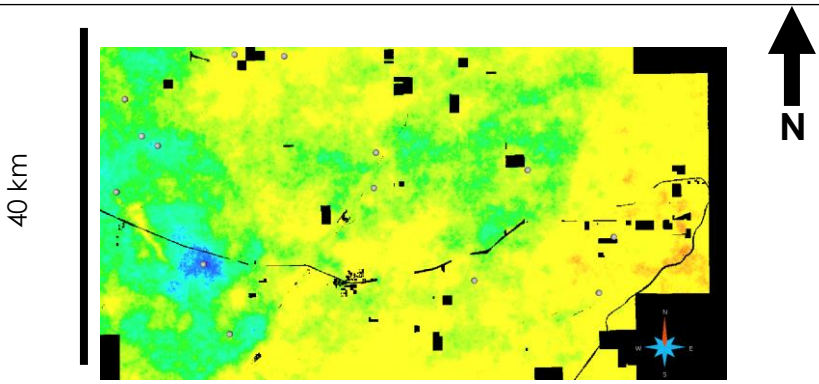


Figure 31: SGS Cokriging, Realization 30 (Time Slice)


1.6----------1.9

Color Bar for Figures 29 through 34 (Vp/Vs, unitless)

Once all realizations have been computed and examined, statistical information can now be derived from the multiple datasets. The simplest of the statistical operations, and probably the most obvious to use is the mean of the realizations:

$$m = \frac{1}{n}\sum_{i=1}^{n} z_i \tag{5.1}$$

The mean of the realizations is displayed in figure 33 below. One may notice the mean is much smoother than each of the unique solutions. It is most definitely smoother than the non-linear regression output, soft data, shown in figure 32. The cokriging output is less erratic compared to the non-linear regression output. This comes with the nature of a smoothed output. The cokriging output is very similar to the control result in figure 34. The largest difference between the two is in the northeast section of the dataset. The next step is to figure out how different the prediction is from the dataset, and how well we can predict this uncertainty without knowing the answer.

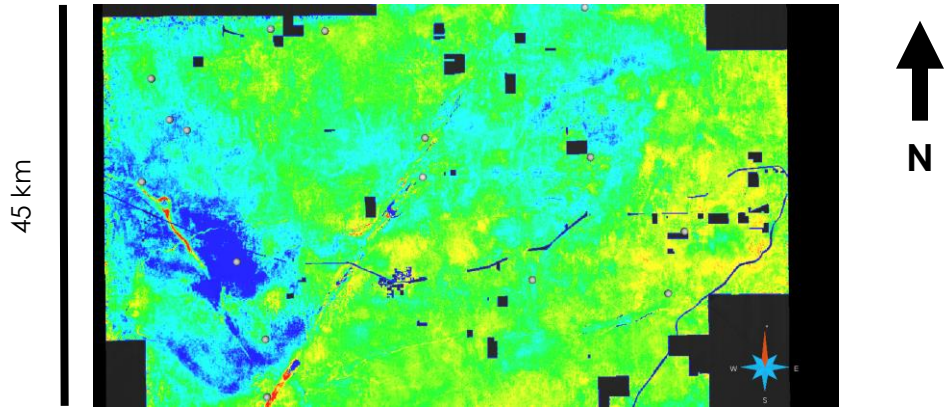


Figure 32:Multi-Attribute Non-Linear Regression (Time Slice)

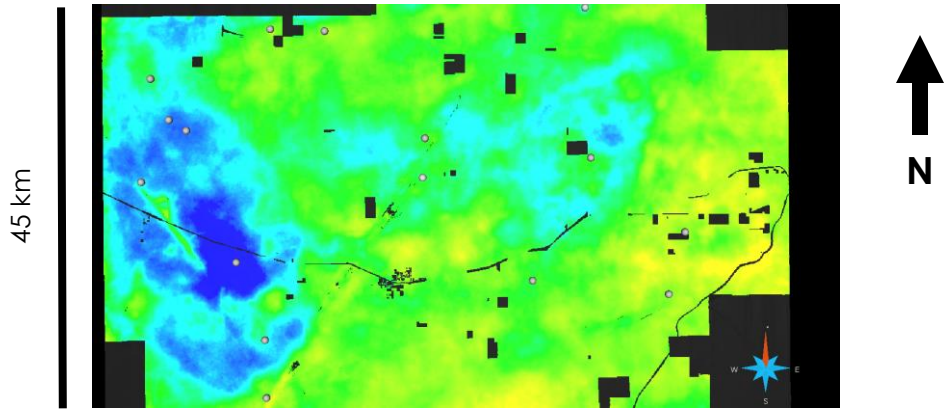


Figure 33: Cokriging Mean of 30 Realizations (Time Slice)

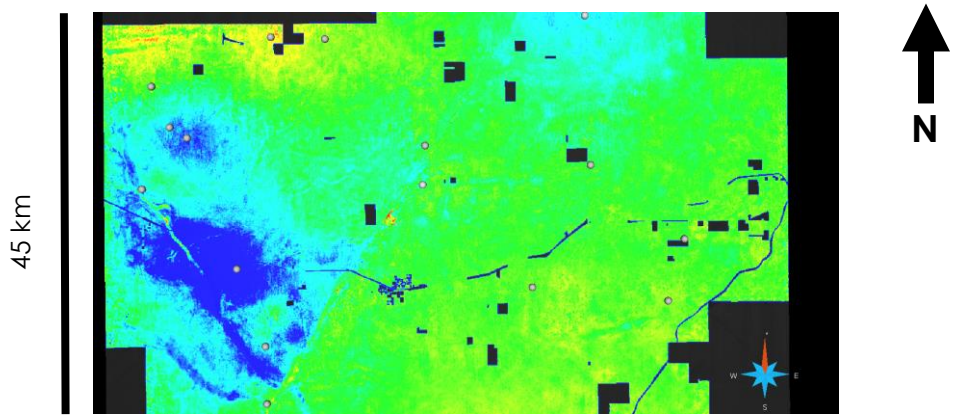


Figure 34: Control Vp/Vs (Time Slice)

5.3 Results and Uncertainty Analysis

Now that the Synthetic prediction has been completed, we can compare the prediction to the control result. Figure 36 below is a map of the absolute value of the difference between the control dataset and predicted dataset. At this scale there are a lot of differences scattered throughout the dataset. Some trends can be seen though, such as that in the south eastern side of the dataset which doesn't show much error. To predict the spatial uncertainty of the SGS Cokriging technique a variance was calculated from all 30 solutions:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - m)^2 \tag{5.2}$$

The variance volume shows in which areas things have changed throughout different realizations. The unit used to measure variance is not at the same scale as the Vp/Vs volume however. Thus we calculate the standard deviation from the variance to make a better comparison

$$\sigma = \sqrt{\sigma^2} \tag{5.3}$$

Figure 35 is the output standard deviation map of the 30 realizations from the cokriging sequential Gaussian simulation. The standard deviation map is much smoother than the control difference map. The standard deviation map is also not showing any high error values.

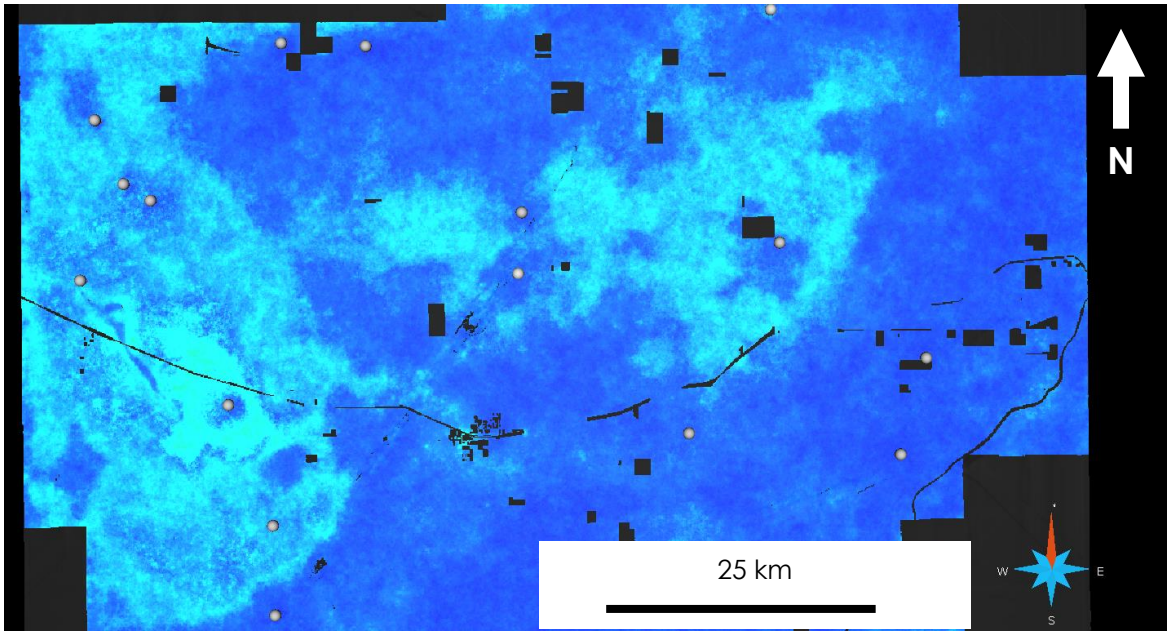


Figure 35: Standard Deviation of 30 Realizations (Time Slice)

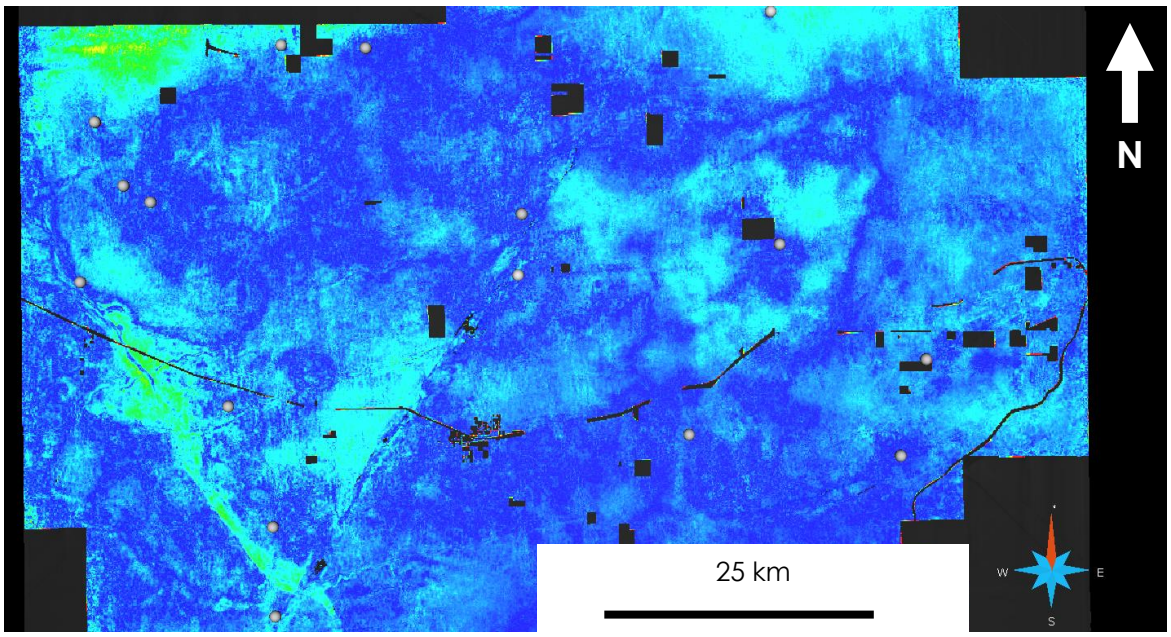
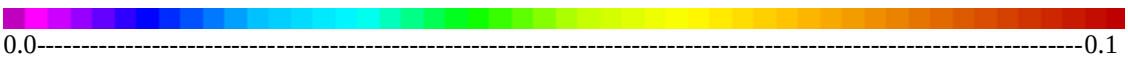


Figure 36: The Absolute Value of the Difference of the Control and Cokriging Vp/Vs (Time Slice)



Color Bar for Figures 35 and 36 (Vp/Vs, Unitless)

Chapter 6 Buffalohorn Dataset Application

6.1 Analysis of Buffalohorn Dataset

Proper Analysis of the Mississippi limestone is the goal of the Buffalohorn dataset. There are multiple properties to predict for that would help in the analysis of the Mississippi Limestone. Some of these properties include mineralogy, porosity, and density. Prediction of mineralogy properties would include silica content, shale content, and calcite content. This may show any small sandstone lenses in the Mississippi Limestone. It could also show where the shale boundary is between the Mississippi Limestone and Woodford Shale. Then again with higher porosity predictions within the Mississippi Limestone may also relate to sandstone lens. Each of these predictions has been created. However for this project the log property to predict for is porosity. The target log is effective porosity. Porosity is defined as the pore volume of the rock divided by the bulk volume.

$$\phi = \frac{Pore\ Volume}{Bulk\ Volume}$$

Effective porosity is defined as the total porosity without any other constituents, examples being shale or clay. The main porosity is the Mississippi Limestone in the region is very small, averaging 3 percent. A neutron log and density log was measured at all well log locations. However, only 14 of the wells will be used in the real dataset. The external attributes that will be used for the Non-Linear Regression will come from a PP-PS joint inversion. The joint inversion was done previously by Shihong Chi at Ion Geophysical. The external attributes include acoustic impedance, shear impedance, and density.

The synthetic dataset in the previous chapters was used to test the process of creating an uncertainty volume using geostatistics and non-linear regression. The real dataset will be using the same process except the answer is not known. The real dataset will also be at a much finer resolution than the synthetic dataset. The output of the non-linear regression will still be based at the seismic sample rate of two milliseconds. Yet the grid in the cokriging will be measured at every 2 feet.

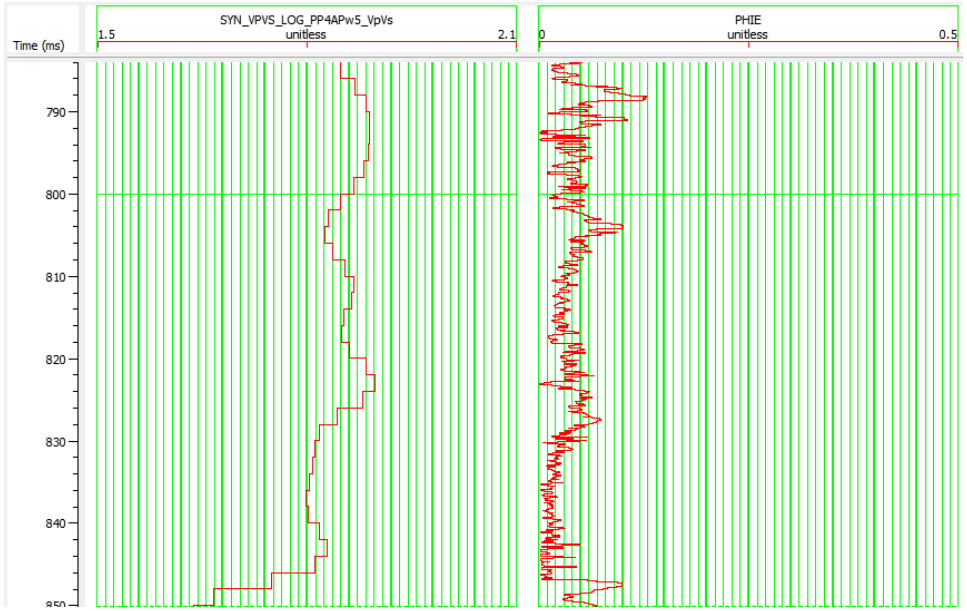


Figure 37: Synthetic Vp/Vs Log Compared to Effective Porosity Log

6.2 Prediction of Log Property using Non-Linear Regression

Similar to the synthetic dataset the first step is to compare a single attribute to the target log of choice. Here the target log, effective porosity, is blocked so that it appears at the same scale as the seismic resolution. Notice that out of all the external attributes the

density log relates the best with the target log. The density attribute is inversely related to the porosity log. This is good because density and porosity are very much related to one another in the real world.

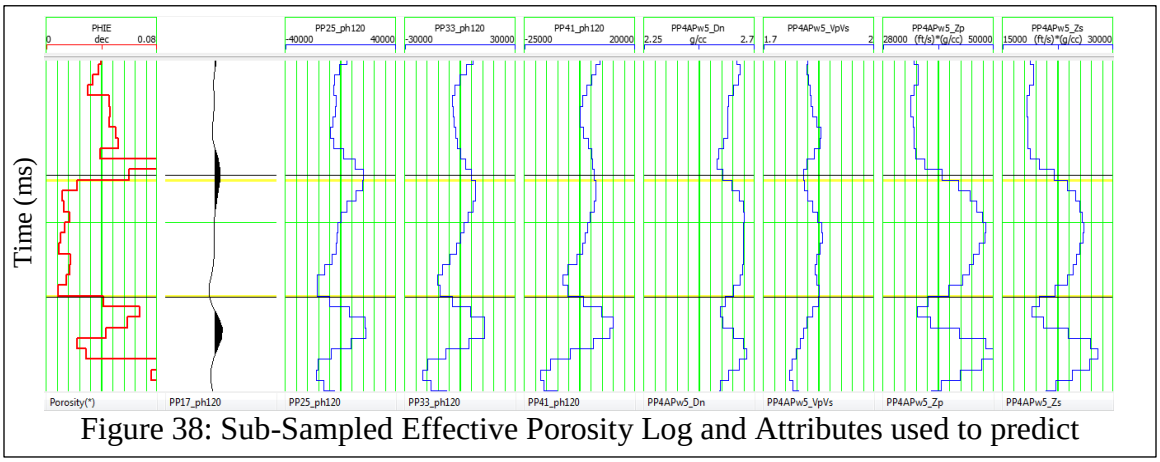


Figure 38: Sub-Sampled Effective Porosity Log and Attributes used to predict

When predicting for the best single attribute different transformations of the density were the highest ranked. The same was true for the multi attribute result. Table 4 below shows the best subset of attributes. The only other external attribute that is in the subset is calculated shear impedance volume. The rest of the attributes were derived from the full angle stack, and alternative angle stacks.

Table 4: Multi-Attribute List for Real Dataset

Rank	Target	Attribute	Training Error	Validation Error
1	ϕ	$\int Density$	0.019234	0.020586
2	ϕ	$SI\ Filter\{25{,}30 \mid 35{,}40\}$	0.018700	0.019995
3	ϕ	$Derivate\ Instantaneous\ Amplitude(40\ Angle\ Stack)$	0.017817	0.019513
4	ϕ	Instantaneous Frequency	0.016860	0.019015
5	ϕ	$Density\ Filter\ \{25{,}30 \mid 35{,}40\}$	0.016246	0.018608

To predict the best subset of attributes stepwise regression was used. However instead of searching for the attribute with the least prediction error, the process searched for the attribute with the lowest validation error. This makes the process much longer to run, but should give us the best result. Figure 39 below shows a graph of the prediction and validation error vs. the number of attributes. One will notice that a sixth attribute may have decreased the prediction error, yet we will move along with five attributes.

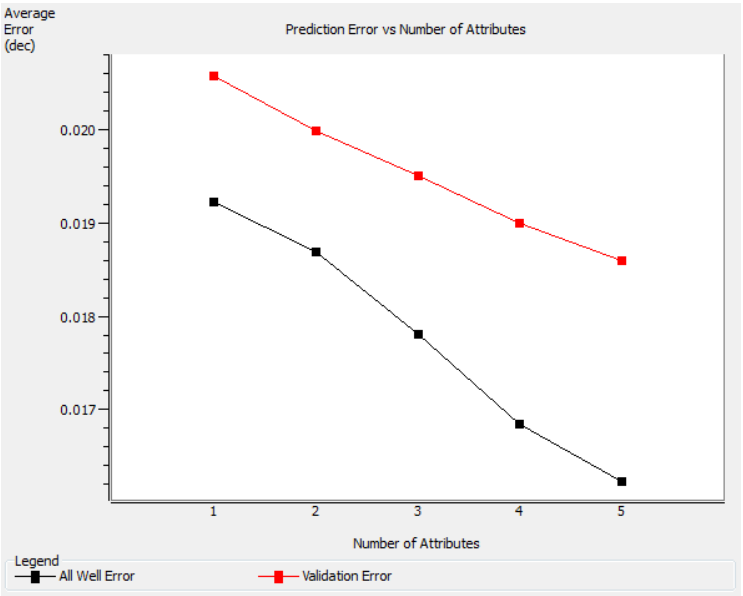


Figure 39: Prediction Error vs Number of Attributes

Now that the best subset of attributes has been identified we need to verify how it looks on the actual prediction. Figure 40 includes three well logs that show validation result using the attributes from table 4. The log is only predicting in the area where the Mississippi limestone was originally picked. The results are definitely not as promising as the synthetic dataset, but for the purpose of predicting the spatial uncertainty this will be valid enough.

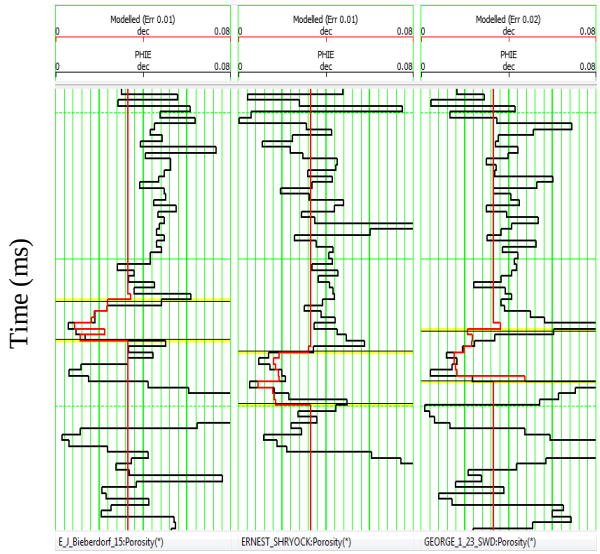


Figure 40: Example of Final Prediction at 3 Well Locations, Black is the Target Log, Red is the Predicted Log

To properly examine how well the prediction is doing a crossplot is made between the prediction and actual porosity. Figure 41 is showing that the porosity is fitting the regression line better on the lower side rather than the higher side. This is because most of the sample points from all the porosity logs are on the low end.

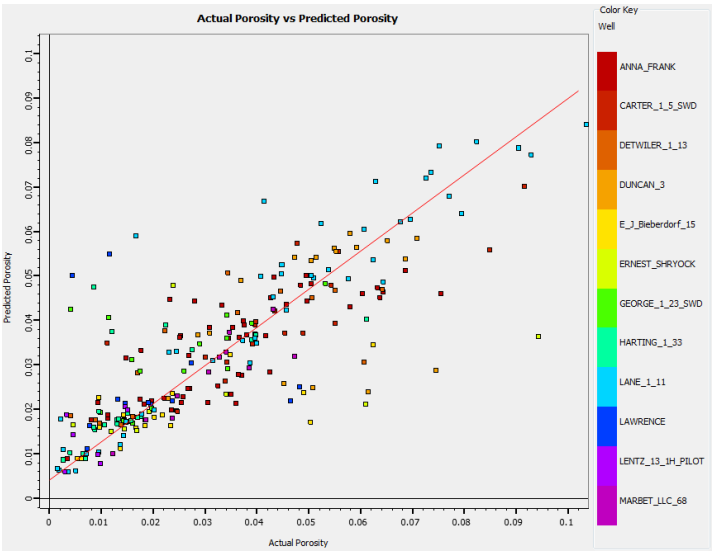


Figure 41: Crossplot of Actual vs Predicted Porosity, Colors are coded per well

The prediction was made using 14 wells out of the 17 wells inside of the seismic dataset. Figure 42 shows a data slice of the output effective porosity volume following the Mississippi horizon. One may notice that in comparison to the synthetic data on figure 19, this data looks a lot more erratic. This is because there is noise in the real seismic compared to the synthetic. In analyzing the porosity slice it seems the central eastern side of the data has a higher porosity than the western side. Figure 43 is a cross section between two wells to see how the data is interpolating. The interpolation does not contain a continuous porosity between both of the wells. Yet this is what someone would expect in a Limestone. The middle of the Mississippi Limestone has higher porosity than the lower and higher sections.

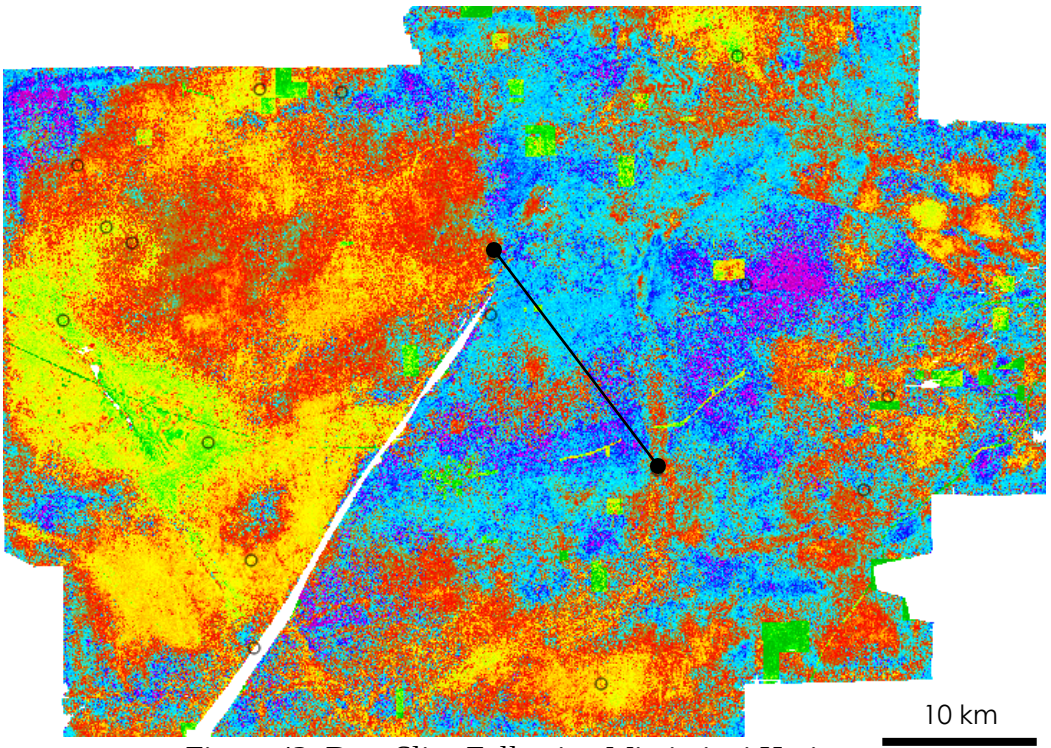


Figure 42: Data Slice Following Mississippi Horizon

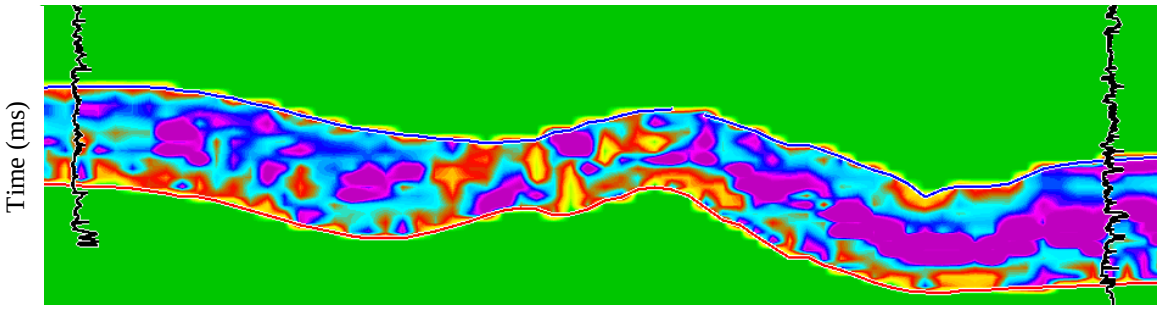


Figure 43: Cross Section of Predicted Porosity, Line A Represented in Figure 42

Color Bar for Figures 42 and 43 (Vp/Vs, Colorless)

6.3 Prediction of Log Property using Geostatistics

The prediction using multi-attribute linear regression was in the time domain at a sample rate of two milliseconds. When predicting for the reservoir property using geostatistics the domain will still be in time, but the sampling will be much finer. The spatial sampling will stay the same, sampling every common depth point of the stacked volume. It is more common place in geostatistical methods to convert to depth before predicting for such a volume, but to not include an additional variable it was decided against. The synthetic dataset had a total number of 17,000,000 cells and was on the same order as the seismic data. The grid for the real dataset has 74,000,000 cells. This increases the computational time significantly for running the real dataset. Yet the increase in vertical resolution helps to keep the character of the well log around the well locations.

The spatial variability between of the hard data, porosity logs, can be views in figures 44 through 46. Each variogram has a different way of telling the story of spatial continuity. The modeled vertical variogram in figure 44 shows a great fit to the data. The areal and omnidirectional variograms show the data has a lot of spatial variability.

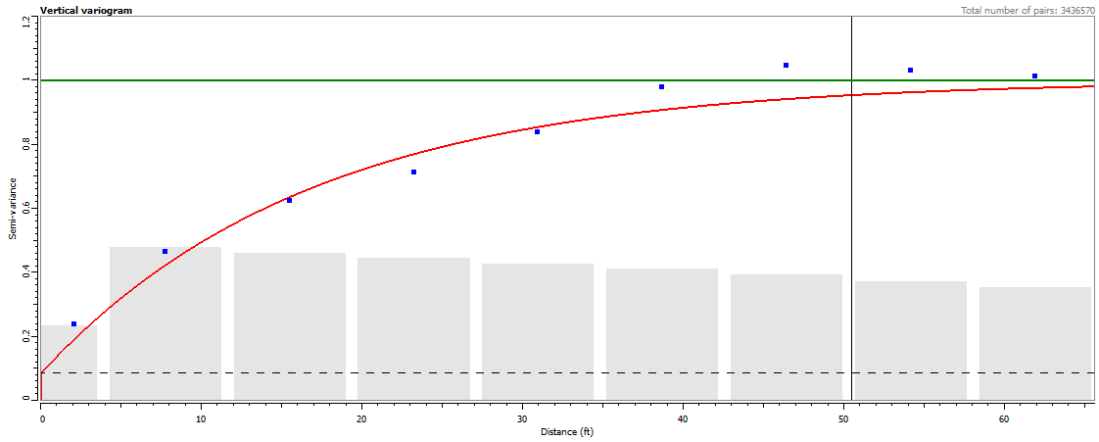


Figure 44: Vertical Variogram

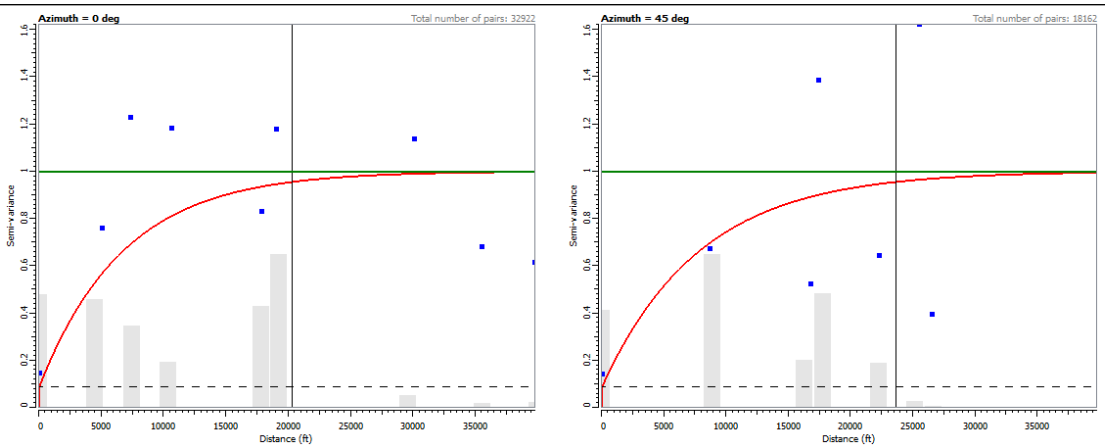


Figure 45: Areal variogram separated by azimuth

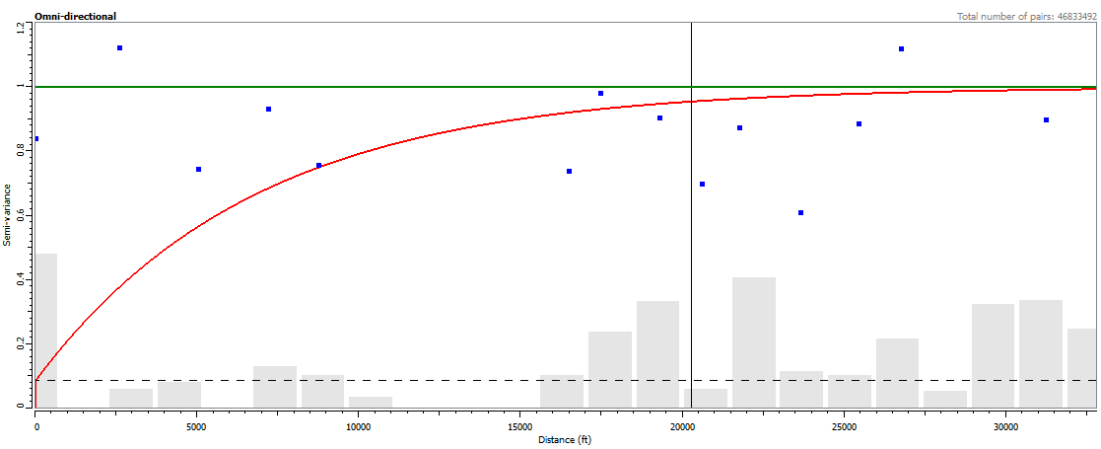


Figure 46: Omnidirectional Variogram

The input histogram for the cokriging will be created from the hard data. The unsmoothed histogram of the effective porosity logs is shown in figure 47. It is easy to notice there is a lot of erratic jumps and noise in the histogram. Since the input histogram is used in the Gaussian simulation to sample from, it would be good if the histogram was smoother and had a Gaussian shape. The synthetic dataset did not have this issue since the Vp/Vs histogram was already relatively bell shaped. The porosity logs however are very close to 0 so its mirror as a half of a Gaussian shape. Figure 48 is a smoothed version of this histogram. Both were tested, and the output of the smoothed histogram proved to be a lot less noisy.

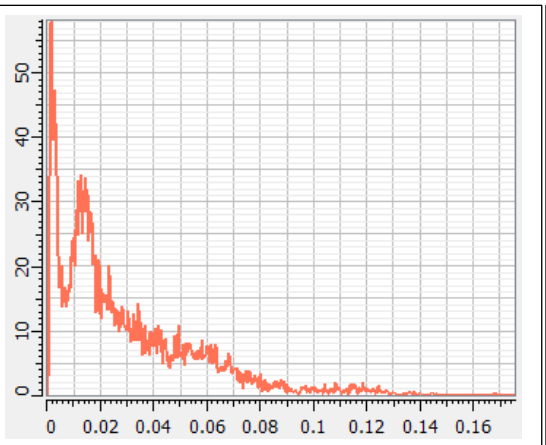


Figure 47: Histogram of Porosity Logs

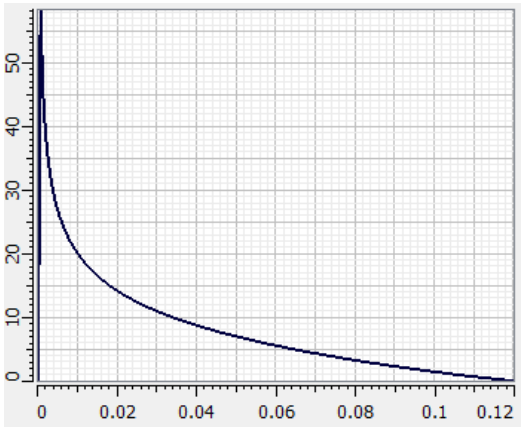


Figure 48: Smoothed Histogram of Porosity Logs

Thirty realizations were run through the simulation. The correlation coefficient of the hard data and the soft data is a lot less than that of the synthetic dataset. The correlation coefficient is a .77 compared to the synthetic which was a .92. Thus there is more trust to the hard data in the range of the variogram modeled in previous sections.

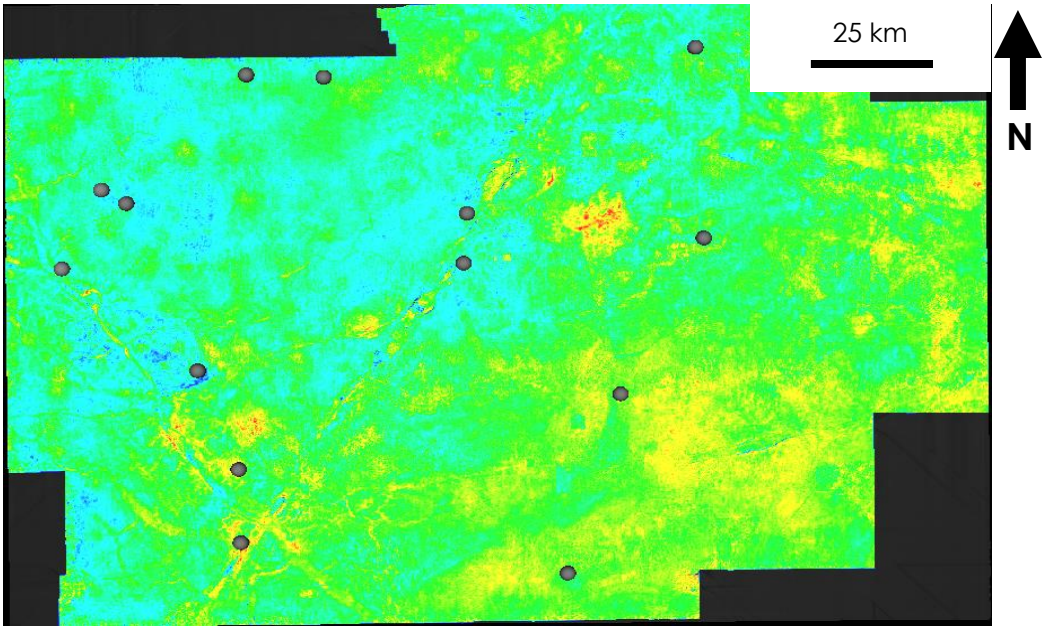


Figure 49: Multi-Attribute Non-Linear Regression Porosity Prediction

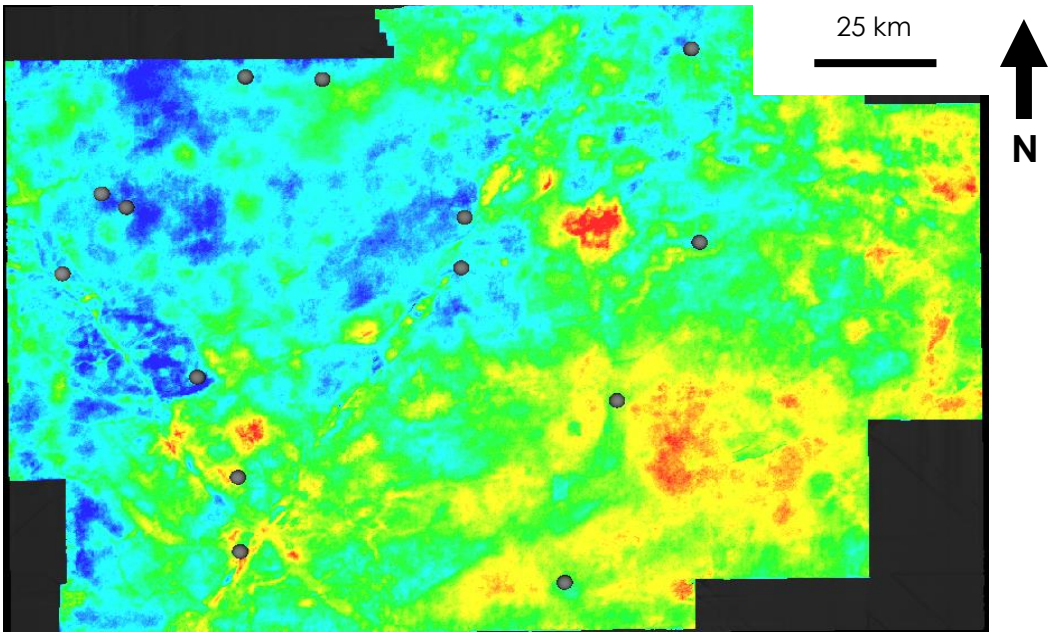
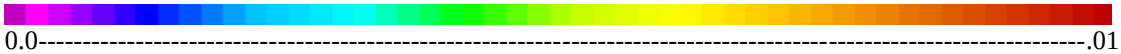


Figure 50: SGS Cokriging Mean Porosity Prediction



Color Bar for Figures 49 and 50 (Predicted porosity, percent)

Figure 50 is an average of all of the simulations and figure 49 is the soft data input, the non-linear regression output. The multi attribute non-linear regression output is more irregular when compared to the simulation output. Figures 51 and 52 are showing a 3D view of the outputs of the comparable processes with two well logs as well.

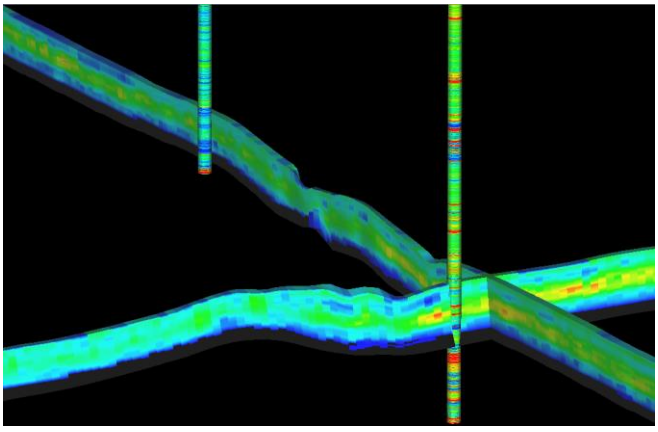


Figure 51: Wells with Multi-Attribute Non-Linear Regression Porosity Prediction (3D View of two cross section lines)

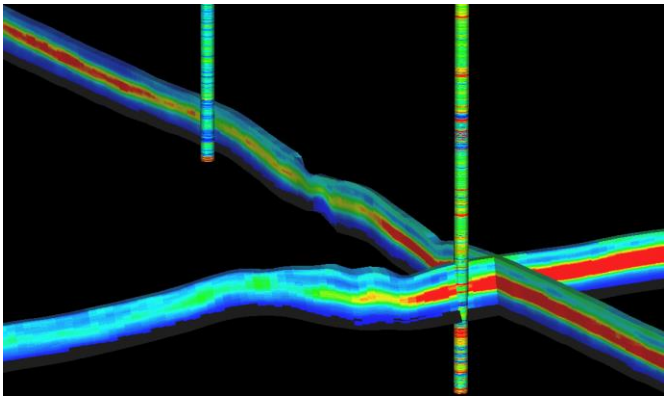


Figure 52: Wells with SGS Cokriging Porosity Prediction (3D View of two cross section lines)

Color Bar for Figures 51 and 52 (Predicted porosity, percent)

In the simulation output it is much easier to differentiate between possible porosity zones in the Mississippi limestone. This is expected because the cokriging application was done on a much finer grid than the multi attribute regression.

6.4 Results and Uncertainty Analysis

The standard deviation volume created from the thirty realizations shows areas of the volume that are weak to predict porosity. In the Comparison of Figure 53 and Figure 50 it is interesting to see that most the areas that have a high standard deviation are in areas of relatively high predicted porosity. An example of this is in the south east section of the volume. In the predicted porosity volume the southeast area has porosity values of up to ten percent. This value of porosity is quite high for the Mississippi Limestone. Similar to the synthetic example the lower error values are being captured, however the higher order of error is not being predicted.

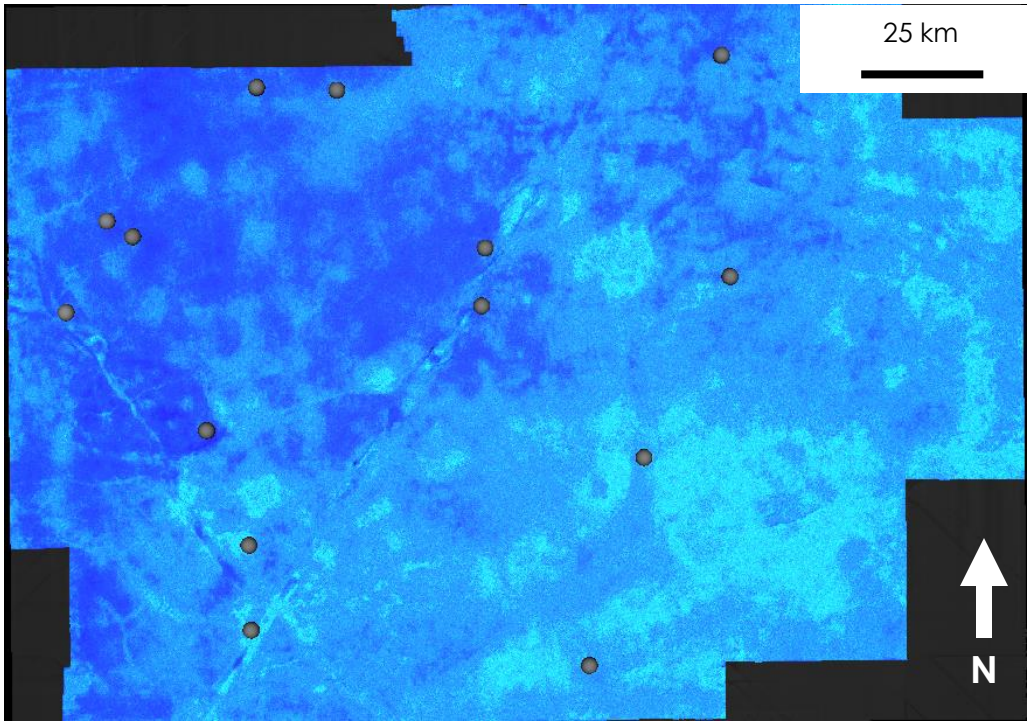


Figure 53: Standard Deviation of 30 Simulations Generated from SGS Cokriging (Time Slice)

Color Bar for Figure 53 (Measurement of Uncertainty)

Chapter 7 Conclusion

7.1 Summary of Results

The use of multi-attribute prediction on the synthetic dataset was able to establish a base framework that can calculate volumetric uncertainty of a reservoir property predicted with geostatistics. The comparison between the uncertainty volume and the actual uncertainty, known from the synthetic dataset, showed that the predicted uncertainty volume was unable to capture the higher values of error in the prediction. This may be due

to the fact that there was such a high correlation between the well data and soft data that the sequential Gaussian simulation did not vary as much as needed. When the framework was applied to the real data a drastic increase in apparent vertical resolution was obtained in the collocated cokriging process due to the grid size. However the uncertainty volume did show similar characteristics to the synthetic volume, only showing areas of low error. Since there is no control uncertainty volume to compare to for the real dataset it remains uncertain how effective the process was. The overall approach shows the potential to calculate volumetric uncertainty in reservoir property volume estimation using 3D seismic data, well logs, and P-wave inversion outputs, which can be computed on a regular basis using multi-attribute linear regression and collocated cokriging.

7.2 Future Work

- A Classification Scheme in the Process of Non-Linear Regression and Cokriging. Most classification has already been done in a vertical sense, the horizons have been picked for the Mississippi Limestone. Yet no classification has been done to show areal regionalization of the dataset. This may help to predict some areas better. An example of this would be the northeast area of the dataset. This area is prone to a lot more structural heterogeneity compared to the rest of the dataset, and only had one well to predict with. It is possible in most software packages to classify this area and well as a different region, and possibly not to predict in this area.

- Instead of Collocated Cokriging the application of Cokriging, or Full Cokriging, may increase the reliability of the uncertainty calculation. Full Cokriging has the ability to predict with multiple soft datasets. Thus the process would be as follows:

1) Predict Best Subset List of Attributes using Non-Linear Regression, and possibly exhaustive search or stepwise-regression
2) Use each attribute in the subset as a different soft dataset
3) A variogram must be modeled for each soft dataset, and how it compares to the hard data
4) A variogram must also be modeled for the hard dataset itself, and each soft dataset itself
5) Run the Full Cokriging Result, examine the output volume, and the variance volume
6) Run Stochastic Simulation to compare the results of the uncertainty vs. the Collocated Cokriging results.

References

[1] Bohling, G. (2005). *Introduction to Geostatistics and Variogram Analysis.* Kansas Geological Survey. (17, 20)

[2] Chambers, L. R., & Yarus, J. M. (2002). *Quantitative Use of Seismic Attributes for Reservoir Characterization.* Broken Arrow. (15)

[3] Hampson, D., Schuelke, J., & Quirein, J. (2001). *Use of Multi-Attribute Transforms to Predict Log Properties from Seismic Data.* (8,9,10,12,18)

[4] Isaaks, E., & Srivastava, R. (1989). *Applied Geostatistics.* Oxford University Press. (23)

[5] Johnson, K. S. (2008). Geologic History of Oklahoma. *Education Publication*, 3-8. (28,29)

[6] Journel, A. G. (1989). Fundamentals of Geostatistics in Five Lessons. *Short Course in Geology: Volume 8*. (20)

[7] Kalkomey, C. T. (1997). Potential risks when using seismic attributes as predictors of reservoir properties. *The Leading Edge*, 247-251. (1,12)

[8] Olea, R. A. (2009). *A Practical Primer on Geostatistics.* Virginia: USGS. (23)

[9] Paradigm. (2012). *SKUA and GOCAD User Guide.* Houston: Paradigm.

[10] Tordorov, T. I. (2000). *Integration of 3C-3D seismic data and well logs for rock property estimation.* Calgary: University of Calgary. (15-26)

[11] Wickstrom, C. W., & Johnson, C. L. (2012, September 24). Structure and Stratigraphy of the Mississippian System, East of the Nemaha Uplift in Oklahoma. *Search and Discovery Article #10442*. (29)